BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PA 232824

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

April 10, 2000

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***60/144,919***
FILING DATE: ***July 21, 1999***

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

P. SWAIN
Certifying Officer

A/prov

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

DOCKET NUMBER
11757.80USP3

CERTIFICATE UNDER 37 CFR 1.10
"Express Mail" mailing label number: EL435536064US
Date of Deposit: July 21, 1999

I hereby certify that this paper or fee is being deposited with the U.S. Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to Assistant Commissioner for Patents, Washington, D.C. 20231.

By: Hassen Buie
Name: Hassen Buie

REQUEST FOR PROVISIONAL APPLICATION UNDER 37 C.F.R. § 1.53(c)

BOX PROVISIONAL PATENT APPLICATION
Assistant Commissioner for Patents
Washington, DC 20231

Dear Sir:

This is a request for filing a Provisional application for patent under 37 CFR § 1.53(c) entitled NOVEL HUMAN KALLIKREIN-LIKE GENES by the following inventor(s):

Full Name Of Inventor	**Family Name** YOUSEF	**First Given Name** George	**Second Given Name** M.
Residence & Citizenship	**City** Toronto	**State or Foreign Country** Ontario Canada	**Country of Citizenship** Egypt
Post Office Address	**Post Office Address** 90 Gerrard St. West, Apt. 1402	**City** Toronto	**State & Zip Code/Country** Ontario M5G 1J6 Canada
Full Name Of Inventor	**Family Name** DIAMANDIS	**First Given Name** Eleftherios	**Second Given Name** P.
Residence & Citizenship	**City** Toronto	**State or Foreign Country** Ontario Canada	**Country of Citizenship** Canada
Post Office Address	**Post Office Address** 44 Gerrard St. West, Suite 1504	**City** Toronto	**State & Zip Code/Country** Ontario M5G 2K2 Canada
Full Name Of Inventor	**Family Name**	**First Given Name**	**Second Given Name**
Residence & Citizenship	**City**	**State or Foreign Country**	**Country of Citizenship**
Post Office Address	**Post Office Address**	**City**	**State & Zip Code/Country**

1. ☒ Enclosed is the Provisional application for patent as follows: 64 pages of specification, and 32 sheets of drawings.

2. ☒ A Verified Statement that this filing is by a small entity (37 CFR 1.9, 1.27, 1.28) is attached.

3. ☒ Payment of Provisional filing fee under 37 C.F.R. § 1.16(k) :
   - ☒ Attached is a check in the amount of $ 75.00.
   - ☐ Please charge Deposit Account No. 13-2725.
   - ☐ PAYMENT OF THE FILING FEE IS BEING DEFERRED.

4. ☒ The Commissioner is hereby authorized to charge any additional fees as set forth in 37 CFR §§ 1.16 to 1.18 which may be required by this paper or credit any overpayment to Account No. 13-2725.

5. ☐ Enclosed is an Assignment of the invention to , Recordation Form Cover Sheet and a check for $ to cover the Recordation Fee.

6. ☐ Also Enclosed:

7. ☐ The invention was made by the following agency of the United States Government or under a contract with the following agency of the United States Government:

8. ☒ Address all future communications to the **Attention of Douglas P. Mueller** (may only be completed by attorney or agent of record) at the address below.

9. ☒ A return postcard is enclosed.

Respectfully submitted,

YOUSEF et al.

By their Attorneys,

MERCHANT, GOULD, SMITH, EDELL,
WELTER & SCHMIDT, P.A.
3100 Norwest Center
90 South Seventh Street
Minneapolis, Minnesota 55402
(612) 332-5300

By
Douglas P. Mueller
Reg No. 30,300
DPM:vvh

Dated: July 21, 1999

MSH File: K[illegible]LIKREIN

UNITED STATES PROVISIONAL II

Title: **Novel Human Kallikrein-Like Genes**

Inventors: **George M. Yousef and Eleftherios P. Diamandis**

MSH File : KALLIKREIN

TITLE: Novel Human Kallikrein-Like Genes

FIELD OF THE INVENTION

The invention relates to nucleic acid molecules, proteins encoded by such nucleic acid molecules; and use of the proteins and nucleic acid molecules

BACKGROUND OF THE INVENTION

Kallikreins and kallikrein-like proteins are a subgroup of the serine protease enzyme family and exhibit a high degree of substrate specificity (1). The biological role of these kallikreins is the selective cleavage of specific polypeptide precursors (substrates) to release peptides with potent biological activity (2). In mouse and rat, kallikreins are encoded by large multigene families. In the mouse genome, at least 24 genes have been identified (3). Expression of 11 of these genes has been confirmed; the rest are presumed to be pseudogenes (4). A similar family of 15-20 kallikreins has been found in the rat genome (5) where at least 4 of these are known to be expressed (6).

Three human kallikrein genes have been described, i.e. prostatic specific antigen (PSA or KLK3) (7), human glandular kallikrein (KLK2) (8) and tissue (pancreatic-renal) kallikrein (KLK1) (9). The PSA gene spans 5.8 Kb of sequence which has been published (7); the KLK2 gene has a size of 5.2 Kb and its complete structure has also been elucidated (8). The KLK1 gene is approximately 4.5 Kb long and the exon sequences and the exon/intron junctions of this gene have been determined (9).

The mouse kallikrein genes are clustered in groups of up to 11 genes on chromosome 7 and the distance between the genes in the various clusters can be as small as 3-7 Kb (3). All three human kallikrein genes have been assigned to chromosome 19q13.2 – 19q13.4 and the distance between PSA and KLK2 has been estimated to be 12 Kb (9).

A major difference between mouse and human kallikreins is that two of the human kallikreins (KLK2 and KLK3) are expressed almost exclusively in the prostate while in animals none of the kallikreins is localized in this organ. Other candidate new members of the human kallikrein gene family include protease M (10) (also named Zyme (11) or neurosin (12) and the normal epithelial cell-specific gene-1 (NES1) (13). Both genes have been assigned to chromosome 19q13.3 (10,14) and show structural homology with other serine proteases and the kallikrein gene family (10-14).

SUMMARY OF THE INVENTION

In efforts to precisely define the relative genomic location of PSA, KLK2, Zyme and NES1 genes, an area spanning approximately 300 Kb of contiguous sequence on human chromosome 19 (19q13.3 –q13.4) was examined. The present inventors were able to identify the relative location of the known kallikrein genes and, in addition, they identified other kallikrein- like genes which exhibit both location proximity and structural similarity with the known members of the human kallikrein family. The novel genes exhibit homology with the currently known members of the kallikrein family and they are co-localized in the same genomic region. These new genes, like the already known kallikreins have utility in various cancers including those of the breast, testicular, and prostate.

The kallikrein-like proteins described herein are individually referred to as "KLK-L1 to KLK-L6", and collectively as "kallikrein-like proteins" or "KLK-L Proteins". The genes encoding the proteins are referred to as "*klk-l1* to *klk-l6*", "kallikrein-like genes" or "*klk-l* genes".

Broadly stated the present invention relates to an isolated nucleic acid molecule which comprises:

(i) a nucleic acid sequence encoding a protein having substantial sequence identity preferably at least 60% sequence identity, with an amino acid sequence of KLK-L1 to KLK-L6 as shown in Tables 2 to 6 or Figure 18;

(ii) a nucleic acid sequence encoding a protein comprising with an amino acid sequence of KLK-L1 to KLK-L6 as shown in Tables 2 to 6 or Figure 18;

(iii) nucleic acid sequences complementary to (i);

(iv) a degenerate form of a nucleic acid sequence of (i);

(v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);

(vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising an amino acid sequence of KLK-L1 to KLK-L6 as shown in Tables 2 to 6 or Figure 18; or

(vii) a fragment, or allelic or species variation of (i), (ii) or (iii).

Preferably, a purified and isolated nucleic acid molecule of the invention comprises:

(i) a nucleic acid sequence comprising the sequence of Figure 2, 3, 4, 5, 6, or 19 wherein

T can also be U;

(ii) nucleic acid sequences complementary to (i), preferably complementary to the full nucleic acid sequence of Figure 2, 3, 4, 5, 6, or 19;

(iii) a nucleic acid capable of hybridizing under stringent conditions to a nucleic acid of (i) or (ii) and preferably having at least 18 nucleotides; or

(iv) a nucleic acid molecule differing from any of the nucleic acids of (i) to (iii) in codon sequences due to the degeneracy of the genetic code.

The invention also contemplates a nucleic acid molecule comprising a sequence encoding a truncation of a KLK-L protein, an analog, or a homolog of a KLK-L Protein or a truncation thereof. (KLK-L Protein and truncations, analogs and homologs of the KLK-L Protein are also collectively referred to herein as "KLK-L Related Proteins").

The nucleic acid molecules of the invention may be inserted into an appropriate expression vector, i.e. a vector that contains the necessary elements for the transcription and translation of the inserted coding sequence. Accordingly, recombinant expression vectors adapted for transformation of a host cell may be constructed which comprise a nucleic acid molecule of the invention and one or more transcription and translation elements linked to the nucleic acid molecule.

The recombinant expression vector can be used to prepare transformed host cells expressing KLK-L Related Proteins. Therefore, the invention further provides host cells containing a recombinant molecule of the invention. The invention also contemplates transgenic non-human mammals whose germ cells and somatic cells contain a recombinant molecule comprising a nucleic acid molecule of the invention, in particular one which encodes an analog of the KLK-L Protein, or a truncation of the KLK-L Protein.

The invention further provides a method for preparing KLK-L Related Proteins utilizing the purified and isolated nucleic acid molecules of the invention. In an embodiment a method for preparing a KLK-L Related Protein is provided comprising (a) transferring a recombinant expression vector of the invention into a host cell; (b) selecting transformed host cells from untransformed host cells; (c) culturing a selected transformed host cell under conditions which allow expression of the KLK-L Related Protein; and (d) isolating the KLK-L Related Protein.

The invention further broadly contemplates an isolated KLK-L Protein comprising an amino acid sequence as shown in Tables 2 to 6, or Figure 18.

The KLK-L Related Proteins of the invention may be conjugated with other molecules, such as proteins, to prepare fusion proteins. This may be accomplished, for example, by the synthesis of N-terminal or C-terminal fusion proteins.

The invention further contemplates antibodies having specificity against an epitope of a KLK-L Related Protein of the invention. Antibodies may be labeled with a detectable substance and used to detect proteins of the invention in tissues and cells.

The invention also permits the construction of nucleotide probes which are unique to the nucleic acid molecules of the invention and/or to proteins of the invention. Therefore, the invention also relates to a probe comprising a nucleic acid sequence of the invention, or a nucleic acid sequence encoding a protein of the invention, or a part thereof. The probe may be labeled, for example, with a detectable substance and it may be used to select from a mixture of nucleotide sequences a nucleic acid molecule of the invention including nucleic acid molecules coding for a protein which displays one or more of the properties of a protein of the invention.

The invention still further provides a method for identifying a substance which binds to a protein of the invention comprising reacting the protein with at least one substance which potentially can bind with the protein, under conditions which permit the formation of complexes between the substance and protein and detecting binding. Binding may be detected by assaying for complexes, for free substance, or for non-complexed protein. The invention also contemplates methods for identifying substances that bind to other intracellular proteins that interact with a KLK-L Related Protein. Methods can also be utilized which identify compounds which bind to KLK-L gene regulatory sequences (e.g. promoter sequences).

Still further the invention provides a method for evaluating a compound for its ability to modulate the biological activity of a KLK-L Related Protein of the invention. For example a substance which inhibits or enhances the interaction of the protein and a substance which binds to the protein may be evaluated. In an embodiment, the method comprises providing a known concentration of a KLK-L Related Protein, with a substance which binds to the protein and a test compound under conditions which permit the formation of complexes between the substance and protein, and removing and/or detecting complexes.

Compounds which modulate the biological activity of a protein of the invention may also be identified using the methods of the invention by comparing the pattern and level of

expression of the protein of the invention in tissues and cells, in the presence, and in the absence of the compounds.

The proteins of the invention and substances and compounds identified using the methods of the invention, and peptides of the invention may be used to modulate the biological activity of a KLK-L Related Protein of the invention, and they may be used in the treatment of conditions such as cancer (e.g. breast, testicular, and prostate cancer). Accordingly, the substances and compounds may be formulated into compositions for administration to individuals suffering from cancer.

Therefore, the present invention also relates to a composition comprising one or more of a protein of the invention, a peptide of the invention, or a substance or compound identified using the methods of the invention, and a pharmaceutically acceptable carrier, excipient or diluent. A method for treating or preventing cancer is also provided comprising administering to a patient in need thereof, a KLK-L Related Protein of the invention, or a composition of the invention.

The present inventors have also identified a novel gene homologous to myelin associated protein designated UG. Therefore the invention provides an isolated nucleic acid molecule which comprises:

(i) a nucleic acid sequence encoding a protein having substantial sequence identity preferably at least 60% sequence identity, with an amino acid sequence as shown in Table 7;

(ii) a nucleic acid sequence encoding a protein comprising with an amino acid sequence of as shown in Table 7;

(iii) nucleic acid sequences complementary to (i);

(iv) a degenerate form of a nucleic acid sequence of (i);

(v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);

(vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising with an amino acid sequence of as shown in Table 7; or

(vii) a fragment, or allelic or species variation of (i), (ii) or (iii).

The invention further contemplates an isolated UG Protein comprising an amino acid

sequence as shown in Table 7.

The general description herein relating to the klk-l nucleic acid molecules, and KLK-L Proteins and KLK-L Related Proteins, antibodies, methods, and compositions are applicable to the novel UG protein and nucleic acid molecule.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

Figure 1 shows an approximate 300 Kb of contiguous genomic sequence around chromosome 19q13.3 - q13.4 represented by 8 contigs, each one shown with its length in Kb. The contig numbers refer to those reported in the Lawrence Livermore National Laboratory website. Note the localization of the seven known genes (PSA, KLK2, Zyme, NES1, HSCCE, neuropsin and TLSP) (see abbreviations for full names of these genes). All genes are represented with arrows denoting the direction of transcription. The gene with no homology to human kallikreins is termed UG (unknown gene). The five new kallikrein-like genes (KLK-L1 to KLK-L5) were numbered from the most centromeric to the most telomeric. Numbers just below or just above the arrows indicate appropriate Kb lengths in each contig. The length of each of these genes may change in the future since not all exons were identified for each new gene, as shown in Tables 2-7.

Figure 2 shows the nucleic acid sequence of KLK-L1;

Figure 3 shows the nucleic acid sequence of KLK-L2;

Figure 4 shows the nucleic acid sequence of KLK-L3;

Figure 5 shows the nucleic acid sequence of KLK-L4;

Figure 6 shows the nucleic acid sequence of KLK-L5;

Figure 7 shows a contiguous genomic sequence around chromosome 19q13.3- q13.4. Genes are represented by horizontal arrows denoting the direction of the coding sequence. Distances between genes are in base pairs.

Figure 8 shows tissue expression of the prostase/KLK-L1 gene as determined by RT-PCR. Actin and PSA are control genes. Interpretations are presented in Table 11.

Figure 9 shows the sequence of PCR product obtained with cDNA from female breast tissue using prostase/KLK-L1 primers. Primer sequences are underlined. The sequence is identical to the sequence obtained from prostatic tissue.

Figure 10 is a blot showing the results of experiments for hormonal regulation of the prostase/KLK-L1 gene in the BT-474 breast carcinoma cell lines. DHT = dihydrotestosterone. Steroids were added at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), pS2 (up-regulated by estrogens) and PSA (up-regulated by androgens and progestins), are control genes. Prostase/KLK-L1 is up-regulated by androgens and progestins.

Figure 11 is a schematic diagram showing comparison of the genomic structure of PSA, KLK1, KLK2, zyme, neuropsin and prostase/KLK-L1 genes. Exons are shown by open boxes and introns by the connecting lines. Arrow head shows the start codons and the vertical arrow represents stop codons. Letters above boxes indicate relative positions of the catalytic triad; H denotes histidine, D aspartic acid and S serine. Roman numbers indicate intron phases. The intron phase refers to the location of the intron within the codon; I denotes that the intron occurs after the first nucleotide of the codon, II the intron occurs after the second nucleotide, 0 the intron occurs between codons. Numbers inside boxes indicate exon lengths in base pairs.

Figure 12 shows the genomic organization and partial genomic sequence of the KLK-L2 gene. Intronic sequences are not shown except for the splice junctions. Introns are shown with lower case letters and exons with capital letters. The start and stop codons are encircled and the exon –intron junctions are boxed. The translated amino acids of the coding region are shown underneath by a single letter abbreviation. The catalytic residues are inside triangles. Putative polyadenylation signal is underlined.

Figure 13 shows an approximate 300 Kb region of almost contiguous genomic sequence around chromosome 19q13.3- q13.4.Genes are represented by horizontal arrows denoting the direction of the coding sequence. Distances between genes are mentioned in base pairs .

Figure 14 shows the alignment of the deduced amino acid sequence of KLK-L2 with members of the kallikrein multi-gene family. Genes are (from top to bottom) : Prostase/KLK-L1, enamel matrix serine proteinase 1 (EMSP1) (GenBank accession # NP_004908), KLK-L2, zyme (GenBank accession # Q92876), neuropsin (GenBank accession # BAA28673), trypsin-

like serine protease (TLSP) (GenBank accession # BAA33404), PSA (GenBank accession # P07288), KLK2 (GenBank accession # P20151), KLK1 (GenBank accession # NP_002248), and trypsinogen (GenBank accession # P07477). Dashes represent gaps to bring the sequences to better alignment. The residues of the catalytic triad are represented by (✤) and the 29 invariant serine protease residues by (❙ or ✤). Conserved areas around the catalytic triad are boxed. The predicted cleavage sites are indicated by (♠). The dotted area represents the kallikrein loop sequence. The trypsin like cleavage pattern is indicated by (✪).

Figure 15A shows a dendrogram of the predicted phylogenetic tree for some kallikrein genes. Neighbor-joining/UPGMA method was used to align KLK-L2 with other members of the kallikrein gene family. Gene names and accession numbers are listed in Figure 14. The tree grouped the classical kallikreins (KLK1, KLK2, and PSA) together and aligned the KLK-L2 gene in one group with EMSP, prostase, and TLSP.

Figure 15B is a plot of hydrophobicity and hydrophilicity of KLK-L2.

Figure 16 is a blot showing tissue expression of KLK-L2 gene as determined by RT-PCR. Actin and PSA are control genes. Interpretations are presented in Table 14.

Figure 17 are blots showing hormonal regulation of the KLK-L2 gene in BT-474 breast carcinoma cell lines. DHT = dihydrotestosterone. Steroids were at 10^{-8} M final concentrations. Actin (not regulated by steroid hormones), pS2 (up-regulated by estrogens) and PSA (upregulated by androgens and progestins), are control genes. KLK-L2 is upregulated by estrogens and progestins.

Figure 18 shows the amino acid sequence of human KLK-L6;

Figure 19 shows the nucleic acid sequence of the gene encoding KLK-L6;

Figure 20 is a schematic diagram showing the kallikrein gene locus.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See for example, Sambrook, Fritsch, & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y); DNA Cloning: A Practical Approach, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization B.D. Hames & S.J. Higgins eds. (1985); Transcription and Translation B.D.

Hames & S.J. Higgins eds (1984); Animal Cell Culture R.I. Freshney, ed. (1986); Immobilized Cells and enzymes IRL Press, (1986); and B. Perbal, A Practical Guide to Molecular Cloning (1984).

1. Nucleic Acid Molecules of the Invention

As hereinbefore mentioned, the invention provides an isolated nucleic acid molecule having a sequence encoding a KLK-L Protein. The term "isolated" refers to a nucleic acid substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical reactants, or other chemicals when chemically synthesized. An "isolated" nucleic acid may also be free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid molecule) from which the nucleic acid is derived. The term "nucleic acid" is intended to include DNA and RNA and can be either double stranded or single stranded. In an embodiment, a nucleic acid molecule encodes a KLK-L Protein comprising an amino acid sequence as shown in Tables 2 to 6 or Figure 18, preferably a nucleic acid molecule comprising a nucleic acid sequence as shown in Figure 2, 3, 4, 5, 6, or Figure 19.

The invention includes nucleic acid sequences complementary to a nucleic acid encoding a KLK-L Protein comprising an amino acid sequence as shown in Tables 2 to 6, preferably the nucleic acid sequences complementary to a full nucleic acid sequence shown in Figure 2, 3, 4, 5, 6, or 19.

The invention includes nucleic acid molecules having substantial sequence identity or homology to nucleic acid sequences of the invention or encoding proteins having substantial identity or similarity to the amino acid sequence shown in Tables 2 to 9, or Figure 18. Preferably, the nucleic acids have substantial sequence identity for example at least 40% nucleic acid identity; more preferably 50% nucleic acid identity; and most preferably at least 60% to 80% sequence identity. "Identity" as known in the art and used herein, is a relationship between two or more amino acid sequences or two or more nucleic acid sequences, as determined by comparing the sequences. It also refers to the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences. Identity and similarity are well known terms to skilled artisans and they can be calculated by conventional methods (for example see Computational Molecular Biology, Lesk, A.M. ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and

Genome Projects, Smith, D.W. ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M. and Griffin, H.G. eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G. Acadmeic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J. eds. M. Stockton Press, New York, 1991, Carillo, H. and Lipman, D., SIAM J. Applied Math. 48:1073, 1988). Methods which are designed to give the largest match between the sequences are generally preferred. Methods to determine identity and similarity are codified in publicly available computer programs including the GCG program package (Devereux J. et al., Nucleic Acids Research 12(1): 387, 1984); BLASTP, BLASTN, and FASTA (Atschul, S.F. et al. J. Molec. Biol. 215: 403-410, 1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S. et al. NCBI NLM NIH Bethesda, Md. 20894; Altschul, S. et al. J. Mol. Biol. 215: 403-410, 1990).

Isolated nucleic acid molecules encoding a KLK-L Protein, and having a sequence which differs from a nucleic acid sequence of the invention due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent proteins (e.g., a KLK-L Protein) but differ in sequence from the sequence of a KLK-L Protein due to degeneracy in the genetic code. As one example, DNA sequence polymorphisms within the nucleotide sequence of a KLK-L Protein may result in silent mutations which do not affect the amino acid sequence. Variations in one or more nucleotides may exist among individuals within a population due to natural allelic variation. Any and all such nucleic acid variations are within the scope of the invention. DNA sequence polymorphisms may also occur which lead to changes in the amino acid sequence of a KLK-L Protein. These amino acid polymorphisms are also within the scope of the present invention.

Another aspect of the invention provides a nucleic acid molecule which hybridizes under stringent conditions, preferably high stringency conditions to a nucleic acid molecule which comprises a sequence which encodes a KLK-L Protein having an amino acid sequence shown in Tables 2 to 6, or Figure 18. Appropriate stringency conditions which promote DNA hybridization are known to those skilled in the art, or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C may be employed. The stringency may be selected based on the conditions used in the wash step. By

way of example, the salt concentration in the wash step can be selected from a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be at high stringency conditions, at about 65°C.

It will be appreciated that the invention includes nucleic acid molecules encoding a KLK-L Related Protein including truncations of a KLK-L Protein, and analogs of a KLK-L Protein as described herein. It will further be appreciated that variant forms of the nucleic acid molecules of the invention which arise by alternative splicing of an mRNA corresponding to a cDNA of the invention are encompassed by the invention.

An isolated nucleic acid molecule of the invention which comprises DNA can be isolated by preparing a labelled nucleic acid probe based on all or part of a nucleic acid sequence of the invention. The labeled nucleic acid probe is used to screen an appropriate DNA library (e.g. a cDNA or genomic DNA library). For example, a cDNA library can be used to isolate a cDNA encoding a KLK-L Related Protein by screening the library with the labeled probe using standard techniques. Alternatively, a genomic DNA library can be similarly screened to isolate a genomic clone encompassing a gene encoding a KLK-L Related Protein. Nucleic acids isolated by screening of a cDNA or genomic DNA library can be sequenced by standard techniques.

An isolated nucleic acid molecule of the invention which is DNA can also be isolated by selectively amplifying a nucleic acid encoding a KLK-L Related Protein using the polymerase chain reaction (PCR) methods and cDNA or genomic DNA. It is possible to design synthetic oligonucleotide primers from the nucleotide sequence of the invention for use in PCR. A nucleic acid can be amplified from cDNA or genomic DNA using these oligonucleotide primers and standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. cDNA may be prepared from mRNA, by isolating total cellular mRNA by a variety of techniques, for example, by using the guanidinium-thiocyanate extraction procedure of Chirgwin et al., Biochemistry, 18, 5294-5299 (1979). cDNA is then synthesized from the mRNA using reverse transcriptase (for example, Moloney MLV reverse transcriptase available from Gibco/BRL, Bethesda, MD, or AMV reverse transcriptase available from Seikagaku America, Inc., St. Petersburg, FL).

An isolated nucleic acid molecule of the invention which is RNA can be isolated by cloning a cDNA encoding a KLK-L Related Protein into an appropriate vector which allows

for transcription of the cDNA to produce an RNA molecule which encodes a KLK-L Related Protein. For example, a cDNA can be cloned downstream of a bacteriophage promoter, (e.g. a T7 promoter) in a vector, cDNA can be transcribed *in vitro* with T7 polymerase, and the resultant RNA can be isolated by conventional techniques.

Nucleic acid molecules of the invention may be chemically synthesized using standard techniques. Methods of chemically synthesizing polydeoxynucleotides are known, including but not limited to solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071).

Determination of whether a particular nucleic acid molecule encodes a KLK-L Related Protein can be accomplished by expressing the cDNA in an appropriate host cell by standard techniques, and testing the expressed protein in the methods described herein. A cDNA encoding a KLK-L Related Protein can be sequenced by standard techniques, such as dideoxynucleotide chain-termination or Maxam-Gilbert chemical sequencing, to determine the nucleic acid sequence and the predicted amino acid sequence of the encoded protein.

The initiation codon and untranslated sequences of a KLK-L Related Protein may be determined using computer software designed for the purpose, such as PC/Gene (IntelliGenetics Inc., Calif.). The intron-exon structure and the transcription regulatory sequences of a gene encoding a KLK-L Related Protein may be confirmed by using a nucleic acid molecule of the invention encoding a KLK-L Related Protein to probe a genomic DNA clone library. Regulatory elements can be identified using standard techniques. The function of the elements can be confirmed by using these elements to express a reporter gene such as the lacZ gene which is operatively linked to the elements. These constructs may be introduced into cultured cells using conventional procedures or into non-human transgenic animal models. In addition to identifying regulatory elements in DNA, such constructs may also be used to identify nuclear proteins interacting with the elements, using techniques known in the art.

In a particular embodiment of the invention, the nucleic acid molecules isolated using the methods described herein are mutant *klk-l* gene alleles. The mutant alleles may be isolated from individuals either known or proposed to have a genotype which contributes to the symptoms of cancer (e.g. breast, testicular, or prostate cancer). Mutant alleles and mutant allele

products may be used in therapeutic and diagnostic methods described herein. For example, a cDNA of a mutant *klk-l* gene may be isolated using PCR as described herein, and the DNA sequence of the mutant allele may be compared to the normal allele to ascertain the mutation(s) responsible for the loss or alteration of function of the mutant gene product. A genomic library can also be constructed using DNA from an individual suspected of or known to carry a mutant allele, or a cDNA library can be constructed using RNA from tissue known, or suspected to express the mutant allele. A nucleic acid encoding a normal *klk-l* gene or any suitable fragment thereof, may then be labeled and used as a probe to identify the corresponding mutant allele in such libraries. Clones containing mutant sequences can be purified and subjected to sequence analysis. In addition, an expression library can be constructed using cDNA from RNA isolated from a tissue of an individual known or suspected to express a mutant *klk-l* allele. Gene products made by the putatively mutant tissue may be expressed and screened, for example using antibodies specific for a KLK-L Related Protein as described herein. Library clones identified using the antibodies can be purified and subjected to sequence analysis.

The sequence of a nucleic acid molecule of the invention, or a fragment of the molecule, may be inverted relative to its normal presentation for transcription to produce an antisense nucleic acid molecule. An antisense nucleic acid molecule may be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art.

2. Proteins of the Invention

An amino acid sequence of a KLK-L Protein comprises a sequence as shown in Tables 2 to 6, or Figure 18.

In addition to proteins comprising an amino acid sequence as shown Tables 2 to 6 or Figure 18 the proteins of the present invention include truncations of a KLK-L Protein, analogs of a KLK-L Protein, and proteins having sequence identity or similarity to a KLK-L Protein, and truncations thereof as described herein (i.e. KLK-L Related Proteins). Truncated proteins may comprise peptides of between 3 and 70 amino acid residues, ranging in size from a tripeptide to a 70 mer polypeptide.

The truncated proteins may have an amino group (-NH2), a hydrophobic group (for example, carbobenzoxyl, dansyl, or T-butyloxycarbonyl), an acetyl group, a 9-fluorenylmethoxy-carbonyl (PMOC) group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the amino terminal end. The

truncated proteins may have a carboxyl group, an amido group, a T-butyloxycarbonyl group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the carboxy terminal end.

The proteins of the invention may also include analogs of a KLK-L Protein, and/or truncations thereof as described herein, which may include, but are not limited to a KLK-L Protein, containing one or more amino acid substitutions, insertions, and/or deletions. Amino acid substitutions may be of a conserved or non-conserved nature. Conserved amino acid substitutions involve replacing one or more amino acids of a KLK-L Protein amino acid sequence with amino acids of similar charge, size, and/or hydrophobicity characteristics. When only conserved substitutions are made the resulting analog is preferably functionally equivalent to a KLK-L Protein. Non-conserved substitutions involve replacing one or more amino acids of the KLK-L Protein amino acid sequence with one or more amino acids which possess dissimilar charge, size, and/or hydrophobicity characteristics.

One or more amino acid insertions may be introduced into a KLK-L Protein. Amino acid insertions may consist of single amino acid residues or sequential amino acids ranging from 2 to 15 amino acids in length.

Deletions may consist of the removal of one or more amino acids, or discrete portions from a KLK-L Protein sequence. The deleted amino acids may or may not be contiguous. The lower limit length of the resulting analog with a deletion mutation is about 10 amino acids, preferably 20 to 40 amino acids.

The proteins of the invention include proteins with sequence identity or similarity to a KLK-L Protein and/or truncations thereof as described herein. Such KLK-L Proteins include proteins whose amino acid sequences are comprised of the amino acid sequences of KLK-L Protein regions from other species that hybridize under selected hybridization conditions (see discussion of stringent hybridization conditions herein) with a probe used to obtain a KLK-L Protein. These proteins will generally have the same regions which are characteristic of a KLK-L Protein. Preferably a protein will have substantial sequence identity for example, about 50% identity, preferably 70 to 80% identity, more preferably at least 90% to 95% identity, and most preferably 98% identity with an amino acid sequence shown in Tables 2 to 6 or Figure 18.

A percent amino acid sequence homology, similarity or identity is calculated as the percentage of aligned amino acids that match the reference sequence using known methods as

described herein.

The invention also contemplates isoforms of the proteins of the invention. An isoform contains the same number and kinds of amino acids as a protein of the invention, but the isoform has a different molecular structure. Isoforms contemplated by the present invention preferably have the same properties as a protein of the invention as described herein.

The present invention also includes KLK-L Related Proteins conjugated with a selected protein, or a marker protein (see below) to produce fusion proteins. Additionally, immunogenic portions of a KLK-L Protein and a KLK-L Protein Related Protein are within the scope of the invention.

A KLK-L Related Protein of the invention may be prepared using recombinant DNA methods. Accordingly, the nucleic acid molecules of the present invention having a sequence which encodes a KLK-L Related Protein of the invention may be incorporated in a known manner into an appropriate expression vector which ensures good expression of the protein. Possible expression vectors include but are not limited to cosmids, plasmids, or modified viruses (e.g. replication defective retroviruses, adenoviruses and adeno-associated viruses), so long as the vector is compatible with the host cell used.

The invention therefore contemplates a recombinant expression vector of the invention containing a nucleic acid molecule of the invention, and the necessary regulatory sequences for the transcription and translation of the inserted protein-sequence. Suitable regulatory sequences may be derived from a variety of sources, including bacterial, fungal, viral, mammalian, or insect genes (For example, see the regulatory sequences described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Selection of appropriate regulatory sequences is dependent on the host cell chosen as discussed below, and may be readily accomplished by one of ordinary skill in the art. The necessary regulatory sequences may be supplied by the native KLK-L Protein and/or its flanking regions.

The invention further provides a recombinant expression vector comprising a DNA nucleic acid molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is linked to a regulatory sequence in a manner which allows for expression, by transcription of the DNA molecule, of an RNA molecule which is antisense to the nucleic acid sequence of a protein of the invention or a fragment thereof. Regulatory sequences linked to the antisense nucleic acid can be chosen which direct the

continuous expression of the antisense RNA molecule in a variety of cell types, for instance a viral promoter and/or enhancer, or regulatory sequences can be chosen which direct tissue or cell type specific expression of antisense RNA.

The recombinant expression vectors of the invention may also contain a marker gene which facilitates the selection of host cells transformed or transfected with a recombinant molecule of the invention. Examples of marker genes are genes encoding a protein such as G418 and hygromycin which confer resistance to certain drugs, β-galactosidase, chloramphenicol acetyltransferase, firefly luciferase, or an immunoglobulin or portion thereof such as the Fc portion of an immunoglobulin preferably IgG. The markers can be introduced on a separate vector from the nucleic acid of interest.

The recombinant expression vectors may also contain genes which encode a fusion moiety which provides increased expression of the recombinant protein; increased solubility of the recombinant protein; and aid in the purification of the target recombinant protein by acting as a ligand in affinity purification. For example, a proteolytic cleavage site may be added to the target recombinant protein to allow separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the recombinant protein.

The recombinant expression vectors may be introduced into host cells to produce a transformant host cell. "Transformant host cells" include host cells which have been transformed or transfected with a recombinant expression vector of the invention. The terms "transformed with", "transfected with", "transformation" and "transfection" encompass the introduction of a nucleic acid (e.g. a vector) into a cell by one of many standard techniques. Prokaryotic cells can be transformed with a nucleic acid by, for example, electroporation or calcium-chloride mediated transformation. A nucleic acid can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofectin, electroporation or microinjection. Suitable methods for transforming and transfecting host cells can be found in Sambrook et al (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Suitable host cells include a wide variety of prokaryotic and eukaryotic host cells. For example, the proteins of the invention may be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus), yeast cells or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1991).

A host cell may also be chosen which modulates the expression of an inserted nucleic acid sequence, or modifies (e.g. glycosylation or phosphorylation) and processes (e.g. cleaves) the protein in a desired fashion. Host systems or cell lines may be selected which have specific and characteristic mechanisms for post-translational processing and modification of proteins. For example, eukaryotic host cells including CHO, VERO, BHK, HeLA, COS, MDCK, 293, 3T3, and WI38 may be used. For long-term high-yield stable expression of the protein, cell lines and host systems which stably express the gene product may be engineered.

Host cells and in particular cell lines produced using the methods described herein may be particularly useful in screening and evaluating compounds that modulate the activity of a KLK-L Related Protein.

The proteins of the invention may also be expressed in non-human transgenic animals including but not limited to mice, rats, rabbits, guinea pigs, micro-pigs, goats, sheep, pigs, non-human primates (e.g. baboons, monkeys, and chimpanzees) [see Hammer et al. (Nature 315:680-683, 1985), Palmiter et al. (Science 222:809-814, 1983), Brinster et al. (Proc Natl. Acad. Sci USA 82:44384442, 1985), Palmiter and Brinster (Cell. 41:343-345, 1985) and U.S. Patent No. 4,736,866)]. Procedures known in the art may be used to introduce a nucleic acid molecule of the invention encoding a KLK-L Related Protein into animals to produce the founder lines of transgenic animals. Such procedures include pronuclear microinjection, retrovirus mediated gene transfer into germ lines, gene targeting in embryonic stem cells, electroporation of embryos, and sperm-mediated gene transfer.

The present invention contemplates a transgenic animal that carries the *KLK-L* gene in all their cells, and animals which carry the transgene in some but not all their cells. The transgene may be integrated as a single transgene or in concatamers. The transgene may be selectively introduced into and activated in specific cell types (See for example, Lasko et al, 1992 Proc. Natl. Acad. Sci. USA 89: 6236). The transgene may be integrated into the chromosomal site of the endogenous gene by gene targeting. The transgene may be selectively

introduced into a particular cell type inactivating the endogenous gene in that cell type (See Gu et al Science 265: 103-106).

The expression of a recombinant KLK-L Related Protein in a transgenic animal may be assayed using standard techniques. Initial screening may be conducted by Southern Blot analysis, or PCR methods to analyze whether the transgene has been integrated. The level of mRNA expression in the tissues of transgenic animals may also be assessed using techniques including Northern blot analysis of tissue samples, *in situ* hybridization, and RT-PCR. Tissue may also be evaluated immunocytochemically using antibodies against KLK-L Protein.

Proteins of the invention may also be prepared by chemical synthesis using techniques well known in the chemistry of proteins such as solid phase synthesis (Merrifield, 1964, J. Am. Chem. Assoc. 85:2149-2154) or synthesis in homogenous solution (Houbenweyl, 1987, Methods of Organic Chemistry, ed. E. Wansch, Vol. 15 I and II, Thieme, Stuttgart).

N-terminal or C-terminal fusion proteins comprising a KLK-L Related Protein of the invention conjugated with other molecules, such as proteins, may be prepared by fusing, through recombinant techniques, the N-terminal or C-terminal of a KLK-L Related Protein, and the sequence of a selected protein or marker protein with a desired biological function. The resultant fusion proteins contain KLK-L Protein fused to the selected protein or marker protein as described herein. Examples of proteins which may be used to prepare fusion proteins include immunoglobulins, glutathione-S-transferase (GST), hemagglutinin (HA), and truncated myc.

3. Antibodies

KLK-L Related Proteins of the invention can be used to prepare antibodies specific for the proteins. Antibodies can be prepared which bind a distinct epitope in an unconserved region of the protein. An unconserved region of the protein is one that does not have substantial sequence homology to other proteins. A region from a conserved region such as a well-characterized domain can also be used to prepare an antibody to a conserved region of a KLK-L Related Protein. Antibodies having specificity for a KLK-L Related Protein may also be raised from fusion proteins created by expressing fusion proteins in bacteria as described herein.

The invention can employ intact monoclonal or polyclonal antibodies, and immunologically active fragments (e.g. a Fab, (Fab)$_2$ fragment, or Fab expression library fragments and epitope-binding fragments thereof), an antibody heavy chain, and antibody light chain, a genetically engineered single chain Fv molecule (Ladner et al, U.S. Pat. No. 4,946,778),

or a chimeric antibody, for example, an antibody which contains the binding specificity of a murine antibody, but in which the remaining portions are of human origin. Antibodies including monoclonal and polyclonal antibodies, fragments and chimeras, may be prepared using methods known to those skilled in the art.

4. Applications of the Nucleic Acid Molecules, KLK-L Related Proteins, and Antibodies of the Invention

The nucleic acid molecules, KLK-L Related Proteins, and antibodies of the invention may be used in the prognostic and diagnostic evaluation of cancer (e.g. breast, testicular, and prostate cancer), and the identification of subjects with a predisposition to cancer (Section 4.1.1 and 4.1.2). Methods for detecting nucleic acid molecules and KLK-L Related Proteins of the invention, can be used to monitor cancer by detecting KLK-L Related Proteins and nucleic acid molecules encoding KLK-L Related Proteins. It would also be apparent to one skilled in the art that the methods described herein may be used to study the developmental expression of KLK-L Related Proteins and, accordingly, will provide further insight into the role of KLK-L Related Proteins. The applications of the present invention also include methods for the identification of compounds that modulate the biological activity of *KLK-L* or KLK-L Related Proteins (Section 4.2). The compounds, antibodies etc. may be used for the treatment of cancer (Section 4.3).

4.1 Diagnostic Methods

A variety of methods can be employed for the diagnostic and prognostic evaluation of cancer (e.g. breast, testicular, and prostate cancer), and the identification of subjects with a predisposition to cancer. Such methods may, for example, utilize nucleic acid molecules of the invention, and fragments thereof, and antibodies directed against KLK-L Related Proteins, including peptide fragments. In particular, the nucleic acids and antibodies may be used, for example, for: (1) the detection of the presence of *KLK-L* mutations, or the detection of either over- or under-expression of *KLK-L* mRNA relative to a non-disorder state or the qualitative or quantitative detection of alternatively spliced forms of *KLK-L* transcripts which may correlate with certain conditions or susceptibility toward such conditions; and (2) the detection of either an over- or an under-abundance of KLK-L Related Proteins relative to a non- disorder state or the presence of a modified (e.g., less than full length) KLK-L Protein which correlates with a disorder state, or a progression toward a disorder state.

The methods described herein may be performed by utilizing pre-packaged diagnostic kits comprising at least one specific *KLK-L* nucleic acid or antibody described herein, which may be conveniently used, e.g., in clinical settings, to screen and diagnose patients and to screen and identify those individuals exhibiting a predisposition to developing a disorder.

Nucleic acid-based detection techniques are described, below, in Section 4.1.1. Peptide detection techniques are described, below, in Section 4.1.2. The samples that may be analyzed using the methods of the invention include those which are known or suspected to express *KLK-L* or contain KLK-L Related Proteins. The samples may be derived from a patient or a cell culture, and include but are not limited to biological fluids, tissue extracts, freshly harvested cells, and lysates of cells which have been incubated in cell cultures.

4.1.1 Methods for Detecting Nucleic Acid Molecules of the Invention

The nucleic acid molecules of the invention allow those skilled in the art to construct nucleotide probes for use in the detection of nucleic acid sequences of the invention in samples. Suitable probes include nucleic acid molecules based on nucleic acid sequences encoding at least 5 sequential amino acids from regions of the KLK-L Protein, preferably they comprise 15 to 30 nucleotides. A nucleotide probe may be labeled with a detectable substance such as a radioactive label which provides for an adequate signal and has sufficient half-life such as ^{32}P, ^{3}H, ^{14}C or the like. Other detectable substances which may be used include antigens that are recognized by a specific labeled antibody, fluorescent compounds, enzymes, antibodies specific for a labeled antigen, and luminescent compounds. An appropriate label may be selected having regard to the rate of hybridization and binding of the probe to the nucleotide to be detected and the amount of nucleotide available for hybridization. Labeled probes may be hybridized to nucleic acids on solid supports such as nitrocellulose filters or nylon membranes as generally described in Sambrook et al, 1989, Molecular Cloning, A Laboratory Manual (2nd ed.). The nucleic acid probes may be used to detect genes, preferably in human cells, that encode KLK-L Related Proteins. The nucleotide probes may also be useful in the diagnosis of cancer; in monitoring the progression of cancer; or monitoring a therapeutic treatment.

The probe may be used in hybridization techniques to detect genes that encode KLK-L Related Proteins. The technique generally involves contacting and incubating nucleic acids (e.g. recombinant DNA molecules, cloned genes) obtained from a sample from a patient or other cellular source with a probe of the present invention under conditions favorable for the specific

annealing of the probes to complementary sequences in the nucleic acids. After incubation, the non-annealed nucleic acids are removed, and the presence of nucleic acids that have hybridized to the probe if any are detected.

The detection of nucleic acid molecules of the invention may involve the amplification of specific gene sequences using an amplification method such as PCR, followed by the analysis of the amplified molecules using techniques known to those skilled in the art. Suitable primers can be routinely designed by one of skill in the art.

Genomic DNA may be used in hybridization or amplification assays of biological samples to detect abnormalities involving *klk- l* structure, including point mutations, insertions, deletions, and chromosomal rearrangements. For example, direct sequencing, single stranded conformational polymorphism analyses, heteroduplex analysis, denaturing gradient gel electrophoresis, chemical mismatch cleavage, and oligonucleotide hybridization may be utilized.

Genotyping techniques known to one skilled in the art can be used to type polymorphisms that are in close proximity to the mutations in a *klk-l* gene. The polymorphisms may be used to identify individuals in families that are likely to carry mutations. If a polymorphism exhibits linkage disequalibrium with mutations in a *klk-l* gene, it can also be used to screen for individuals in the general population likely to carry mutations. Polymorphisms which may be used include restriction fragment length polymorphisms (RFLPs), single-base polymorphisms, and simple sequence repeat polymorphisms (SSLPs).

A probe of the invention may be used to directly identify RFLPs. A probe or primer of the invention can additionally be used to isolate genomic clones such as YACs, BACs, PACs, cosmids, phage or plasmids. The DNA in the clones can be screened for SSLPs using hybridization or sequencing procedures.

Hybridization and amplification techniques described herein may be used to assay qualitative and quantitative aspects of *klk-l* expression. For example, RNA may be isolated from a cell type or tissue known to express *klk-l* and tested utilizing the hybridization (e.g. standard Northern analyses) or PCR techniques referred to herein. The techniques may be used to detect differences in transcript size which may be due to normal or abnormal alternative splicing. The techniques may be used to detect quantitative differences between levels of full length and/or alternatively splice transcripts detected in normal individuals relative to those individuals exhibiting cancer symptoms or other disease conditions.

The primers and probes may be used in the above described methods *in situ* i.e directly on tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections.

4.1.2 Methods for Detecting KLK-L Related Proteins

Antibodies specifically reactive with a KLK-L Related Protein, or derivatives, such as enzyme conjugates or labeled derivatives, may be used to detect KLK-L Related Proteins in various samples (e.g. biological materials). They may be used as diagnostic or prognostic reagents and they may be used to detect abnormalities in the level of KLK-L Related Proteins expression, or abnormalities in the structure, and/or temporal, tissue, cellular, or subcellular location of a KLK-L Related Protein. Antibodies may also be used to screen potentially therapeutic compounds *in vitro* to determine their effects on cancer, and other conditions. *In vitro* immunoassays may also be used to assess or monitor the efficacy of particular therapies. The antibodies of the invention may also be used *in vitro* to determine the level of *KLK-L* expression in cells genetically engineered to produce a KLK-L Related Protein.

The antibodies may be used in any known immunoassays which rely on the binding interaction between an antigenic determinant of a KLK-L Related Protein and the antibodies. Examples of such assays are radioimmunoassays, enzyme immunoassays (e.g. ELISA), immunofluorescence, immunoprecipitation, latex agglutination, hemagglutination, and histochemical tests. The antibodies may be used to detect and quantify KLK-L Related Proteins in a sample in order to determine its role in particular cellular events or pathological states, and to diagnose and treat such pathological states.

In particular, the antibodies of the invention may be used in immuno-histochemical analyses, for example, at the cellular and sub-subcellular level, to detect a KLK-L Related Protein, to localize it to particular cells and tissues, and to specific subcellular locations, and to quantitate the level of expression.

Cytochemical techniques known in the art for localizing antigens using light and electron microscopy may be used to detect a KLK-L Related Protein. Generally, an antibody of the invention may be labeled with a detectable substance and a KLK-L Related Protein may be localised in tissues and cells based upon the presence of the detectable substance. Examples of detectable substances include, but are not limited to, the following: radioisotopes (e.g., ^{3}H, ^{14}C, ^{35}S, ^{125}I, ^{131}I), fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), luminescent labels such as luminol; enzymatic labels (e.g., horseradish peroxidase, beta-galactosidase,

luciferase, alkaline phosphatase, acetylcholinesterase), biotinyl groups (which can be detected by marked avidin e.g., streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods), predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, metal binding domains, epitope tags). In some embodiments, labels are attached via spacer arms of various lengths to reduce potential steric hindrance. Antibodies may also be coupled to electron dense substances, such as ferritin or colloidal gold, which are readily visualised by electron microscopy.

The antibody or sample may be immobilized on a carrier or solid support which is capable of immobilizing cells, antibodies etc. For example, the carrier or support may be nitrocellulose, or glass, polyacrylamides, gabbros, and magnetite. The support material may have any possible configuration including spherical (e.g. bead), cylindrical (e.g. inside surface of a test tube or well, or the external surface of a rod), or flat (e.g. sheet, test strip). Indirect methods may also be employed in which the primary antigen-antibody reaction is amplified by the introduction of a second antibody, having specificity for the antibody reactive against KLK-L Related Protein. By way of example, if the antibody having specificity against a KLK-L Related Protein is a rabbit IgG antibody, the second antibody may be goat anti-rabbit gamma-globulin labeled with a detectable substance as described herein.

Where a radioactive label is used as a detectable substance, a KLK-L Related Protein may be localized by radioautography. The results of radioautography may be quantitated by determining the density of particles in the radioautographs by various optical methods, or by counting the grains.

4.2 Methods for Identifying or Evaluating Substances/Compounds

The methods described herein are designed to identify substances that modulate the biological activity of a KLK-L Related Protein including substances that bind to KLK-L Related Proteins, or bind to other proteins that interact with a KLK-L Related Protein, to compounds that interfere with, or enhance the interaction of a KLK-L Related Protein and substances that bind to the KLK-L Related Protein or other proteins that interact with a KLK-L Related Protein. Methods are also utilized that identify compounds that bind to *KLK-L* regulatory sequences.

The substances and compounds identified using the methods of the invention include

but are not limited to peptides such as soluble peptides including Ig-tailed fusion peptides, members of random peptide libraries and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids, phosphopeptides (including members of random or partially degenerate, directed phosphopeptide libraries), antibodies [e.g. polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, single chain antibodies, fragments, (e.g. Fab, $F(ab)_2$, and Fab expression library fragments, and epitope-binding fragments thereof)], and small organic or inorganic molecules. The substance or compound may be an endogenous physiological compound or it may be a natural or synthetic compound.

Substances which modulate a KLK-L Related Protein can be identified based on their ability to bind to a KLK-L Related Protein. Therefore, the invention also provides methods for identifying substances which bind to a KLK-L Related Protein. Substances identified using the methods of the invention may be isolated, cloned and sequenced using conventional techniques.

Substances which can bind with a KLK-L Related Protein may be identified by reacting a KLK-L Related Protein with a test substance which potentially binds to a KLK-L Related Protein, under conditions which permit the formation of substance-KLK-L Related Protein complexes and removing and/or detecting the complexes. The complexes can be detected by assaying for substance-KLK-L Related Protein complexes, for free substance, or for non-complexed KLK-L Related Protein. Conditions which permit the formation of substance-KLK-L Related-Protein complexes may be selected having regard to factors such as the nature and amounts of the substance and the protein.

The substance-protein complex, free substance or non-complexed proteins may be isolated by conventional isolation techniques, for example, salting out, chromatography, electrophoresis, gel filtration, fractionation, absorption, polyacrylamide gel electrophoresis, agglutination, or combinations thereof. To facilitate the assay of the components, antibody against KLK-L Related Protein or the substance, or labeled KLK-L Related Protein, or a labeled substance may be utilized. The antibodies, proteins, or substances may be labeled with a detectable substance as described above.

A KLK-L Related Protein, or the substance used in the method of the invention may be insolubilized. For example, a KLK-L Related Protein, or substance may be bound to a suitable carrier such as agarose, cellulose, dextran, Sephadex, Sepharose, carboxymethyl cellulose polystyrene, filter paper, ion-exchange resin, plastic film, plastic tube, glass beads, polyamine-

methyl vinyl-ether-maleic acid copolymer, amino acid copolymer, ethylene-maleic acid copolymer, nylon, silk, etc. The carrier may be in the shape of, for example, a tube, test plate, beads, disc, sphere etc. The insolubilized protein or substance may be prepared by reacting the material with a suitable insoluble carrier using known chemical or physical methods, for example, cyanogen bromide coupling.

The invention also contemplates a method for evaluating a compound for its ability to modulate the biological activity of a KLK-L Related Protein of the invention, by assaying for an agonist or antagonist (i.e. enhancer or inhibitor) of the binding of a KLK-L Related Protein with a substance which binds with a KLK-L Related Protein. The basic method for evaluating if a compound is an agonist or antagonist of the binding of a KLK-L Related Protein and a substance that binds to the protein, is to prepare a reaction mixture containing the KLK-L Related Protein and the substance under conditions which permit the formation of substance-KLK-L Related Protein complexes, in the presence of a test compound. The test compound may be initially added to the mixture, or may be added subsequent to the addition of the KLK-L Related Protein and substance. Control reaction mixtures without the test compound or with a placebo are also prepared. The formation of complexes is detected and the formation of complexes in the control reaction but not in the reaction mixture indicates that the test compound interferes with the interaction of the KLK-L Related Protein and substance. The reactions may be carried out in the liquid phase or the KLK-L Related Protein, substance, or test compound may be immobilized as described herein. The ability of a compound to modulate the biological activity of a KLK-L Related Protein of the invention may be tested by determining the biological effects on cells.

It will be understood that the agonists and antagonists i.e. inhibitors and enhancers that can be assayed using the methods of the invention may act on one or more of the binding sites on the protein or substance including agonist binding sites, competitive antagonist binding sites, non-competitive antagonist binding sites or allosteric sites.

The invention also makes it possible to screen for antagonists that inhibit the effects of an agonist of the interaction of KLK-L Related Protein with a substance which is capable of binding to the KLK-L Related Protein. Thus, the invention may be used to assay for a compound that competes for the same binding site of a KLK-L Related Protein.

The invention also contemplates methods for identifying compounds that bind to

proteins that interact with a KLK-L Related Protein. Protein-protein interactions may be identified using conventional methods such as co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Methods may also be employed that result in the simultaneous identification of genes which encode proteins interacting with a KLK-L Related Protein. These methods include probing expression libraries with labeled KLK-L Related Protein.

Two-hybrid systems may also be used to detect protein interactions *in vivo*. Generally, plasmids are constructed that encode two hybrid proteins. A first hybrid protein consists of the DNA-binding domain of a transcription activator protein fused to a KLK-L Related Protein, and the second hybrid protein consists of the transcription activator protein's activator domain fused to an unknown protein encoded by a cDNA which has been recombined into the plasmid as part of a cDNA library. The plasmids are transformed into a strain of yeast (e.g. *S. cerevisiae*) that contains a reporter gene (e.g. lacZ, luciferase, alkaline phosphatase, horseradish peroxidase) whose regulatory region contains the transcription activator's binding site. The hybrid proteins alone cannot activate the transcription of the reporter gene. However, interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

It will be appreciated that fusion proteins may be used in the above-described methods. In particular, KLK-L Related Proteins fused to a glutathione-S-transferase may be used in the methods.

The reagents suitable for applying the methods of the invention to evaluate compounds that modulate a KLK-L Related Protein may be packaged into convenient kits providing the necessary materials packaged into suitable containers. The kits may also include suitable supports useful in performing the methods of the invention.

4.3 Compositions and Treatments

The proteins of the invention, substances or compounds identified by the methods described herein, antibodies, and antisense nucleic acid molecules of the invention may be used for modulating the biological activity of a KLK-L Related Protein, and they may be used in the treatment of conditions such as cancer (e.g. prostate, testicular, or breast cancer). Accordingly, the substances, antibodies, peptides, and compounds may be formulated into pharmaceutical compositions for administration to subjects in a biologically compatible form suitable for

administration *in vivo*. By "biologically compatible form suitable for administration *in vivo*" is meant a form of the active substance to be administered in which any toxic effects are outweighed by the therapeutic effects. The active substances may be administered to living organisms including humans, and animals. Administration of a therapeutically active amount of a pharmaceutical composition of the present invention is defined as an amount effective, at dosages and for periods of time necessary to achieve the desired result. For example, a therapeutically active amount of a substance may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of antibody to elicit a desired response in the individual. Dosage regima may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation.

The active substance may be administered in a convenient manner such as by injection (subcutaneous, intravenous, etc.), oral administration, inhalation, transdermal application, or rectal administration. Depending on the route of administration, the active substance may be coated in a material to protect the substance from the action of enzymes, acids and other natural conditions that may inactivate the substance.

The compositions described herein can be prepared by per se known methods for the preparation of pharmaceutically acceptable compositions which can be administered to subjects, such that an effective quantity of the active substance is combined in a mixture with a pharmaceutically acceptable vehicle. Suitable vehicles are described, for example, in Remington's Pharmaceutical Sciences (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., USA 1985). On this basis, the compositions include, albeit not exclusively, solutions of the active substances in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and iso-osmotic with the physiological fluids.

Based upon their homology to genes encoding kallikrein, nucleic acid molecules of the invention may be also useful in the treatment of conditions such as hypertension, cardiac hypertrophy, arthritis, inflammatory disorders, neurological disorders, and blood clotting disorders.

Vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, may be used to deliver nucleic acid molecules to a targeted organ,

tissue, or cell population. Methods well known to those skilled in the art may be used to construct recombinant vectors which will express antisense nucleic acid molecules of the invention. (See, for example, the techniques described in Sambrook et al (supra) and Ausubel et al (supra)).

The nucleic acid molecules comprising full length cDNA sequences and/or their regulatory elements enable a skilled artisan to use sequences encoding a protein of the invention as an investigative tool in sense (Youssoufian H and H F Lodish 1993 Mol Cell Biol 13:98-104) or antisense (Eguchi et al (1991) Annu Rev Biochem 60:631-652) regulation of gene function. Such technology is well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the coding or control regions.

Genes encoding a protein of the invention can be turned off by transfecting a cell or tissue with vectors which express high levels of a desired KLK-L-encoding fragment. Such constructs can inundate cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until all copies are disabled by endogenous nucleases.

Modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA or PNA, to the regulatory regions of a gene encoding a protein of the invention, ie, the promoters, enhancers, and introns. Preferably, oligonucleotides are derived from the transcription initiation site, eg, between -10 and +10 regions of the leader sequence. The antisense molecules may also be designed so that they block translation of mRNA by preventing the transcript from binding to ribosomes. Inhibition may also be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Therapeutic advances using triplex DNA were reviewed by Gee J E et al (In: Huber B E and B I Carr (1994) Molecular and Immunologic Approaches, Futura Publishing Co, Mt Kisco N.Y.).

Ribozymes are enzymatic RNA molecules that catalyze the specific cleavage of RNA. Ribozymes act by sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. The invention therefore contemplates engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding a protein of the invention.

Specific ribozyme cleavage sites within any potential RNA target may initially be

identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once the sites are identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be determined by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Methods for introducing vectors into cells or tissues include those methods discussed herein and which are suitable for *in vivo*, *in vitro* and *ex vivo* therapy. For *ex vivo* therapy, vectors may be introduced into stem cells obtained from a patient and clonally propagated for autologous transplant into the same patient (See U.S. Pat. Nos. 5,399,493 and 5,437,994). Delivery by transfection and by liposome are well known in the art.

The nucleic acid molecules disclosed herein may also be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including but not limited to such properties as the triplet genetic code and specific base pair interactions.

The activity of the proteins, substances, compounds, antibodies, nucleic acid molecules, and compositions of the invention may be confirmed in animal experimental model systems.

The following non-limiting examples are illustrative of the present invention:

Examples

Example 1

MATERIALS AND METHODS

Identification of positive PAC and BAC genomic clones from a human genomic DNA library

The sequence of PSA, KLK1, KLK2, NES1 and Zyme genes is already known. Polymerase chain reaction (PCR)-based amplification protocols have been developed which allowed generation of PCR products specific for each one of these genes. Using these PCR products as probes, labeled with ^{32}P, a human genomic DNA PAC library and a human genomic DNA BAC library was screened for the purpose of identifying positive clones of approximately 100-150 Kb long. The general strategies for these experiments have been published elsewhere (14). The genomic libraries were spotted in duplicate on nylon membranes and positive clones

were further confirmed by Southern blot analysis as described (14).

DNA sequences on chromosome 19

The Lawrence Livermore National Laboratory participates in the sequencing of the human genome project and focuses on sequencing chromosome 19. Large sequencing information on this chromosome is available at the website of the Lawrence Livermore National Laboratory (http://www-bio.llnl.gov/genome/gemnome.html).

Approximately 300 Kb of genomic sequences were obtained from that website, encompassing a region on chromosome 19q13.3 - 13.4, where the known kallikrein genes are localized. This 300 Kb of sequence is represented by 8 contigs of variable lengths. By using a number of different computer programs, an almost contiguous sequence of the region was established as shown diagramatically in Figure 1 and Figure 20. Some of the contigs were reversed as shown in Figure 1 in order to reconstruct the area on both strands of DNA.

By using the published sequences of PSA, KLK2, NES1 and Zyme and the computer software BLAST 2, using alignment strategies, the relative positions of these genes on the contiguous map were identified (Figure 1). These known genes served as hallmarks for further studies. An EcoR1 restriction map of the area is also available at the website of the Lawrence Livermore National Laboratory. Using this restriction map and the computer program WebCutter (http://www.firstmarket.com/cutter/cut2.html), a restriction study analysis of the available sequence was performed to further confirm the assignment and relative positions of these contigs along chromosome 19. The obtained configuration and the relative location of the known genes are presented in Figure 1.

Gene prediction analysis

For exon prediction analysis of the whole genomic area, a number of different computer programs were used. These programs are listed in Table 1. All these programs were initially tested using known genomic sequences of the PSA, Zyme, and NES1 genes. The more reliable computer programs, GeneBuilder (gene prediction), GeneBuilder (exon prediction), Grail 2 and GENEID-3 were selected for further use.

Protein homology searching

Putative exons of the new genes were first translated to the corresponding aminoacid sequences. BLAST homology searching for the proteins encoded by the exons of the putative new genes were performed using the BLASTP program and the Genbank databases.

RESULTS

Relative position of PSA, KLK2, Zyme and NES1 on Chromosome 19

Screening of the human BAC library identified two clones which were positive for the Zyme gene (clones BAC 288H1 and BAC 76F7). These BACs were further analyzed by PCR and primers specific for PSA, NES1, KLK1 and KLK2. These analyses indicated that both BACs were positive for Zyme, PSA and KLK2 and negative for KLK1 and NES1 genes.

Screening of the human PAC genomic library identified a PAC clone which was positive for NES1 (clone PAC 34B1). Further PCR analysis indicated that this PAC clone was positive for NES1 and KLK1 genes and negative for PSA, KLK2 and Zyme. Combination of this information with the EcoR1 restriction map of the region allowed establishment of the relative positions of these four genes. PSA is the most centromeric, followed by KLK2, Zyme and NES1. Further alignment of the known sequences of these genes with the 300 Kb contig enabled precise localization of all four genes and determination of the direction of transcription, as shown by the arrows in Figure 1. The KLK1 gene sequence was not identified on any of these contig and appears to be further telomeric to NES1 (since it is co-localized on the same PAC as NES1).

Identification of new genes

A set of rules was used to consider the presence of a new gene in the genomic area of interest as follows:

1. Clusters of at least 3 exons should be found.
2. Only exons with high prediction score ("good" or "excellent" quality, as indicated by the searching programs) were considered for the construction of the putative new genes.
3. Exons predicted were reliable only if they were identified by at least two different exon prediction programs.

By using this strategy, eleven putative new genes were identified of which three were found on subsequent homology analysis to be known genes not previously mapped i.e. the human stratum corneum chymotrypsin enzyme (HSCCE), human neuropsin, and trypsin-like serine protease (TLSP). Their relative location is shown in Figure 1. In addition, one other putative new gene (gene UG) was identified which showed no homology, at the protein level, with the kallikrein proteins. The five remaining genes all have variable homologies with known human or animal kallikrein proteins and/or other known serine proteases (depicted as KLK-L1,

KLK-L2, KLK-L3, KLK-L4 and KLK-L5 in Figure 1 and KLK-L1 to KLK-L6 in Figure 20).

In Tables 2 to 7, the preliminary exon structure and partial protein sequence for each one of the newly identified genes is shown. In Table 8, some proteins are presented which appear, on preliminary analysis, to be homologous to the proteins encoded by the putative new genes. Figure 18 shows the amino acid sequence of KLK-L6 and Figure 19 shows the nucleic acid of the gene encoding KLK-L6.

DISCUSSION

Prediction of protein-coding genes in newly sequenced DNA becomes very important after the establishment of large genome sequencing projects. This problem is complicated due to the exon-intron structure of the eukaryotic genes which interrupts the coding sequence in many unequal parts. In order to predict the protein-coding exons and overall gene structure, a number of computer programs were developed. All these programs are based on the combination of potential functional signals with the global statistical properties of known protein-coding regions (15). However, the most powerful approach for gene structure prediction is to combine information about potential functional signals (splice sites, translation start or stop signal etc.) together with the statistical properties of coding sequences (coding potential) along with information about homologies between the predicted protein and known protein families (16).

In mouse and rat, kallikreins are encoded by large multigene families and these genes tend to cluster in groups with a distance as small as 3.3 – 7.0 Kb (3). A strong conservation of gene order between human chromosome 19q13.1 – q13.4 and 17 loci in a 20-cM proximal part of mouse chromosome 7, including the kallikrein locus, has been documented (17).

In humans, only a few kallikrein genes were identified. In fact, only KLK1, KLK2 and KLK3 (PSA) are considered to represent the human kallikrein gene family (9). The work described herein provides strong evidence that a large number of kallikrein-like genes are clustered within a 300Kb region around chromosome 19q13.2 – q13.4. The three established human kallikreins (KLK1, KLK2, KLK3), Zyme and NES1, as well as the stratum corneum chymotrypticn enzyme, neuropsin, and TLSP (trypsin-like serine protease) and another five new genes, KLK-L1 to KLK-L5, may constitute a large gene family. This will bring the total number of kallikrein or kallikrein-like genes in this region of chromosome 19 to thirteen.

The human stratum corneum chymotryptic enzyme (19), neuropsin (20) and trypsin-like

serine protease (TLSP) (21) are three previously characterized genes which have many structural similarities with the kallikreins and other members of the serine protease family. However, they have not been mapped in the past. Their precise mapping in the region of the kallikrein gene family indicates that these three genes, along with the ones that were newly identified, or are already known, constitute a family that likely originated by duplication of an ancestral gene. The relative localization of all these genes is depicted in Figure 1.

Kallikrein genes are a subfamily of serine proteases, traditionally characterized by their ability to liberate lysyl-bradykinin (kallidin) from kininogen (18). More recently, however, a new, structural concept has emerged to describe kallikreins. From accumulated sequence data, it is now clear that the mouse has many genes with high homology to kallikrein coding sequences (19-20). Richard and co-workers have contributed to the concept of a " kallikrein multigene family" to refer to these genes (21-22). This definition is not based much on specific enzymatic function of the gene product, but more on its sequence homology and their close linkage on mouse chromosome 7. In humans, only KLK1 meets the functional definition of a kallikrein. KLK2 has trypsin-like enzymatic activity and KLK3 (PSA) has very weak chymotrypsin-like enzymatic activity. These activities of KLK2 and KLK3 are not known to liberate biologically active peptides from precursors. Based on the newer definition, members of the kallikrein family include, not only the gene for the kallikrein enzyme, but also genes encoding other homologous proteases, including the enzyme that processes the precursors of the nerve growth factor and epidermal growth factor (8). Therefore, it is important to note the clear distinction between the enzyme kallikrein and a kallikrein or a kallikrein-like gene.

In carrying out the study only exons were considered which were predicted with "good" or "excellent" quality and only exons were considered which were predicted by at least two different programs. Moreover, the presence of a putative gene was only considered when at least three exons clustered coordinately in that region. Additional evidence that these new genes are indeed homologous to the known kallikreins and other serine proteases comes from comparison of the intron phases. As published previously (14), trypsinogen, PSA and NES1 have 5 coding exons of which the first has intron phase I (the intron occurs after the first nucleotide of the codon), the second has intron phase II (the intron occurs after the second nucleotide and the codon), the third has intron phase I and the fourth has intron phase 0 (the intron occurs between codons). The fifth exon contains the stop codon. The intron phases of

the predicted new kallikrein-like genes follow these rules and are shown in the respective tables. Further support comes from the identification in the new genes, of the conserved amino acids of the catalytic domain of the serine proteases, as presented in Tables 2 - 6.

In order to test the accuracy of the computer programs, known genomic areas containing the PSA, Zyme and KLK2 genes were tested. Two of these programs (Grail 2 and GeneBuilder) were able to detect about 95% of the tested known genes. Matches with expressed sequence tag sequences (EST) can also be employed for gene structure prediction in the GeneBuilder program and this can significantly improve the power of the program especially at high stringency (e.g. >95% homology).

In mouse, ten of the kallikrein genes appear to be pseudogenes (9). One of the new genes (UG) does not show homology with the kallikrein genes. However, it has some protein homology with myelin associated glycoprotein (Table 8). There may still be an association between UG and the kallikrein genes since some mouse kallikreins are related to nerve growth factor, as discussed earlier (8) and Zyme as well as neuropsin and TLSP, were found to be highly expressed in brain tissue and it is claimed that Zyme may be related to Alzheimer's disease (11).

Example 2

PROSTASE/KLK-L1 in prostate and breast tissues:

The fine mapping of the prostase/KLK-L1 gene and its chromosomal localization in relation to a number of other homologous genes also mapping to the same region are described. In addition, extensive tissue expression studies were carried out that demonstrate that, in addition to prostate (which shows the highest expression), that prostase/KLK-L1 is also expressed in female breasts, testis, adrenals, uterus, colon, thyroid, brain, spinal cord and salivary glands. Furthermore, the gene is up-regulated by androgens and progestins in the breast carcinoma cell line BT-474.

Materials and Methods

DNA sequences on chromosome 19

Large DNA sequencing data for chromosome 19 is available at the web site of the Lawrence Livermore National Laboratory (LLNL). (http://www-bio.llnl.gov/genome /genome.html). Approximately 300 Kb of genomic sequence was obtained from that web site, encompassing a region on chromosome 19q13.3 - 13.4, where the known kallikrein genes are

localized. This sequence is represented by 9 contigs of variable lengths. By using the sequences of PSA, KLK2, NES1 and protease M and the alignment program BLAST 2, the relative positions of these genes on the contiguous map were located.

Gene prediction analysis

For exon prediction analysis of the whole genomic area, a number of different computer programs were used. Originally all these programs were tested using the known genomic sequences of the PSA, protease M and NES1 genes. The most reliable computer programs GeneBuilder (gene prediction)[http://l25.itba.mi.cnr.it/~webgene/genebuilder.html] GeneBuilder (exon prediction) [http://l25.itba.mi.cnr.it/~webgene/genebuilder.html] , Grail 2 [http://compbio.ornl.gov], and GENEID-3 [http://apolo.imim.es/geneid.htmll] were selected for further use.

Protein homology searching

Putative exons of the newly identified gene were first translated to the corresponding amino acid sequences. BLAST homology searching for the proteins encoded by the exons were performed using the BLASTP program and the GenBank databases.

Searching expressed sequence tags (ESTs)

Sequence homology searching was performed using the BLASTN alogrithm on the National Center for Biotechnology Information web server (http://www ncbi.nlm.nih.gov/BLAST/) against the human EST database (dbEST). Clones with > 95% homology were obtained from the I.M.A.G.E. consortium through Research Genetics Inc, Huntsville, AL and from The Institute for Genomic Research (TIGR) (http://WWW.TIGR.ORG/ tdb/tdb.html) (Table 9). Clones were propagated, purified and then sequenced from both directions with an automated sequencer, using insert-flanking vector primers.

Breast cancer cell line and stimulation experiments

The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (ATCC), Rockville, MD. BT-474 cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. 24 hours before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments, various steroid

hormones dissolved in 100% ethanol were added into the culture media, at a final concentration of 10^{-8} M. Cells stimulated with 100% ethanol were included as controls. The cells were cultured for 24 hours, then harvested for mRNA extraction.

Reverse transcriptase polymerase chain reaction

Total RNA was extracted from the breast cancer cells using Trizol reagent (Gibco BRL) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. 2 µg of total RNA was reverse transcribed into first strand cDNA using the Superscript™ preamplification system (Gibco BRL). The final volume was 20 µl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were designed (Table 10), PCR was carried out in a reaction mixture containing 1 µl of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µl dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ, USA) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 9 minutes to activate the Taq Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 63°C for 1 minute and a final extension at 63°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least 2 exons to avoid contamination by genomic DNA.

Tissue expression of KLK-L1

Total RNA isolated from 26 different human tissues was purchased from Clontech, Palo Alto, CA. cDNA was prepared as described above for the tissue culture experiments and used it for PCR reactions with the primers described in Table 10. Tissue cDNAs were amplified at various dilutions.

Cloning and sequencing of the PCR products

To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, by an automated DNA sequencer.

Results

Identification of the prostase/KLK-L1 gene

The exon prediction strategy of the 300Kb DNA sequences around chromosome 19q13.3

- q13.4 identified a novel gene with a structure reminiscent of a serine protease. The major features of this gene were its homology, at the amino acid and DNA level, with other human kallikrein genes; the conservation of the catalytic triad (histidine, aspartic acid, and serine), the number of exons and the complete conservation of the intron phases.

EST sequence homology search

EST sequence homology search of the putative exons obtained from the gene prediction programs (as described above) against the human EST database (dbEST) revealed five expressed sequence tags (ESTs) with >95 % identity to the putative exons of the gene (Table 9). Positive clones were obtained and the inserts were sequenced from both directions. Alignment was used to compare between the EST sequences and the exons predicted by the computer programs, and final selection of the exon-intron splice sites was made according to the EST sequences. Furthermore, many of the ESTs were overlapping, further ensuring the accuracy of the data.

Mapping and chromosomal localization of prostase /KLK-L1 gene

Alignment of the prostase/ KLK-L1 sequence and the sequences of other known kallikrein genes within the 300 Kb area of the contigs constructed at the Lawrence Livermore National Laboratory enabled precise localization of all genes and to determine the direction of transcription, as shown in Figure 7. The distance between PSA and KLK2 genes was calculated to be 12,508 bp. The prostase/KLK-L1 gene is 26,229 bp more telomeric to KLK2 and transcribes in the opposite direction. The zyme gene is about 51 Kb more telomeric to the prostase gene and transcribes in the same direction. The human stratum corneum chymotryptic enzyme gene, the neuropsin gene and the NES 1 gene are all further telomeric to zyme and all transcribe in the same direction as zyme.

Tissue expression of the prostase/KLK-L1 gene

The tissues that express the prostase/KLK-L1 gene were assessed by RT-PCR. The experiments were performed at various dilutions of the cDNAs to obtain some information about the relative levels of expression. RT-PCR for actin was used as a positive control and RT-PCR for the PSA cDNA was used as another positive control with tissue restricted specificity. Positive ESTs for prostase/KLK-L1 were used as controls for the PCR procedure. The PSA gene was found to be highly expressed in the prostate, as expected, and to a lower extent in mammary and salivary glands as also expected from recent literature reports (24, 25). Very low expression

of PSA in the thyroid gland, trachea and testis was also found, a finding that accords with recent RT-PCR data by others (26).

The tissue expression of prostase/KLK-L1 is summarized in Table 11 and Figure 8. This protease is primarily expressed in the prostate, testis, adrenals, uterus, thyroid, colon, central nervous system and mammary tissues, and, at much lower levels in other tissues. The specificity of the RT-PCR procedure was verified for prostase/KLK-L1 by cloning the PCR products from mammary, testicular and prostate tissues and sequencing them. One example with mammary tissue is shown in Figure 9. All cloned PCR products were identical in sequence to the cDNA sequence reported for the prostase/KLK-L1.

Hormonal regulation of the prostase/KLK-L1 gene

The steroid hormone receptor-positive breast carcinoma cell line BT-474 was used as a model system to evaluate whether prostase/KLK-L1 expression is under steroid hormone regulation. As shown in Figure 10, the controls worked as expected i. e., actin positivity without hormonal regulation in all cDNAs, only estrogen up-regulation of the pS2 gene and up-regulation of the PSA gene by androgens and progestins. Prostase/KLK-L1 is up-regulated primarily by androgens and progestins, similarly to PSA. This up-regulation was dose-dependent and it was evident at steroid hormone levels $\geq 10^{-10}$ M (data not shown).

DISCUSSION

The KLK3 gene encodes for PSA, a protein that currently represents the best tumor marker available (24). Since in rodents there are so many kallikrein genes, the restriction of this family to only 3 genes in humans was somewhat surprising. More recently, new candidate kallikrein genes in humans have been discovered, including NES1 (13) and zyme/protease M/neurosin (10-12). The known kallikreins and the newly discovered kallikrein-like genes share the following similarities: (a) they encode serine proteases (b) they have five coding exons (c) they share significant DNA and protein homologies with each other (d) they map in the same locus on chromosome 19q13.3-q13.4, a region that is structurally similar to an area on mouse chromosome 7, where all the mouse kallikrein genes are localized (e) they appear to be regulated by steroid hormones. Prostase/KLK-L1 is a member of the same family since these common characteristics are also shared by the newly discovered genes.

The exact localization of the KLK-L1 gene and its position in relation to other genes in the area (Figure 7) was determined. Prostase/KLK-L1 lies between KLK2 and zyme.

Irwin et al: (27) have proposed that the serine protease genes can be classified into five different groups according to intron position. The established kallikreins (KLK1, KLK2, and PSA), trypsinogen and chymotrypsinogen belong to a group that has: (1) an intron just downstream from the codon for the active site histidine residue, (2) a second intron downstream from the exon containing the codon for the active site aspartic acid residue, and (3) a third intron just upstream from the exon containing the codon for the active site serine residue. As seen in Figure 11, the genomic organization of prostase/KLK-L1 gene is very similar to this group of genes. The lengths of the coding parts of exons 1-5 are 61,163, 263, 137 and 153 bp, respectively, which are close or identical to the lengths of the exons of the kallikrein genes and also, similar or identical to those of other newly discovered genes in the same chromosomal region like the NES1(14), zyme/protease M/neurosin (10-12) and neuoropsin (28) genes.

The sensitive RT-PCR protocol reveals that the KLK-L1 enzyme is also expressed in significant amounts in other tissues, including testis, female mammary gland, adrenals, uterus, thyroid, colon, brain, lung and salivary glands (Figure 8 and Table 11). The specificity of our RT-PCR primers was verified by sequencing the obtained PCR products, with one example shown in Figure 9. Tissue culture studies with the breast carcinoma cell line BT-474 further confirm not only the ability of these cells to produce prostase/KLK-L1 but also its hormonal regulation (Figure 10).

An interesting theme is now developing involving the group of homologous genes on chromosome 19q13.3(PSA, KLK2, prostase, zyme, and NES1). The combined data suggest that all of them are expressed in prostate and breast tissues, and all of them are hormonally regulated. All these genes may be part of a cascade pathway that plays a role in cell proliferation, differentiation or apoptosis by regulating (positively or negatively) growth factors or their receptors or cytokines, through proteolysis (30). Also interesting is the linkage of locus 19q13 to solid tumors and gliomas (31) which raises the possibility that some of the genes in the region may be disrupted by rearrangements.

The KLK-1L gene encodes for a serine protease that shows homology with other members of the kallikrein gene family and maps to the same chromosomal location. Many structural features of the kallikreins are conserved in prostase/KLK-L1. The precise mapping of this gene between the two known genes KLK2 and zyme is presented. It is further demonstrated that prostase/KLK-L1 is expressed in many tissues, in addition to the prostate,

including the female breast. This gene is also herein referred to as 'prostase'. It has been further demonstrated, using breast carcinoma cell lines, that prostase/KLK-L1 can be produced by these cells and that its expression is significantly up-regulated by androgens and progestins. Based on information for other homologous genes in the area (PSA, zyme, and NES1), prostase/KLK-L1 may be involved in the pathogenesis and/or progression of prostate, breast and possibly other cancers.

Example 3

IDENTIFICATION OF THE KLK-L2 GENE

Materials and Methods

DNA sequence on chromosome 19

Sequencing data of approximately 300Kb of nucleotides on chromosome 19q13.3-q13.4 was obtained from the web site of the Lawrence Livermore National Laboratory (LLNL) (http://www-bio.llnl.gov/genome/genome.html). This sequence was in the form of 9 contigs of different lengths. A restriction analysis study of the available sequences was performed using the "WebCutter" computer program (http://www.firstmarket.com/cutter/cut2.html) and with the aid of the EcoR1 restriction map of this area (also available from the LLNL web site) an almost contiguous stretch of genomic sequences was constructed. The relative positions of the known kallikrein genes: PSA (GenBank accession # X14810), KLK2 (GenBank accession # M18157), and zyme (GenBank accession # U60801) was determined using the alignment program BLAST 2.

NEW GENE IDENTIFICATION

A number of computer programs were used to predict the presence of putative new genes in the genomic area of interest. These programs were initially tested using the known genomic sequences of the PSA, protease M and NES1 genes. The most reliable computer programs GeneBuilder (gene prediction) (http://l25.itba.mi.cnr.it/~webgene/genebuilder.html) GeneBuilder (exon prediction) (http://l25.itba.mi.cnr.it/~webgene/genebuilder.html), Grail 2 (http://compbio.ornl.gov) and GENEID-3 (http://apolo.imim.es/geneid.html) were selected for further use.

Expressed sequence tag (EST) searching

The predicted exons of the putative new gene were subjected to homology search using the BLASTN algorithm on the National Center for Biotechnology Information web server (http://www ncbi.nlm.nih.gov/BLAST/) against the human EST database (dbEST). Clones with > 95% homology were obtained from the I.M.A.G.E. consortium (20) through Research Genetics Inc, Huntsville, AL (Table 12). The clones were propagated, purified and sequenced from both directions with an automated sequencer, using insert-flanking vector primers.

Rapid amplification of cDNA ends (5' RACE)

According to the EST sequence data and the predicted structure of the gene, two gene-specific primers were designed (R1 & R2) (Table 13). Two rounds of RACE reactions (nested PCR) were performed with 5µl Marathon Ready™ cDNA of human testis (Clontech, Palo Alto, CA, USA) as a template. The reaction mix and PCR conditions were conducted according to the manufacturer's recommendations. In brief, denaturation was done for 5 min at 94°C followed by 94° C for 5 sec followed by 72°C for 2 min for 5 cycles, then 94°C for 5 sec followed by 70° C for 2 min for 5 cycles then 94°C for 5 sec followed by 65°C for 2 min for 30 cycles for the first reaction and 25 cycles for the nested PCR reaction.

Tissue expression

Total RNA isolated from 26 different human tissues was purchased from Clontech, Palo Alto, CA. cDNA was prepared as described below for the tissue culture experiments and used for PCR reactions with the primers described in Table 13. Tissue cDNAs were amplified at various dilutions.

Breast cancer cell line and hormonal stimulation experiments

The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (ATCC), Rockville, MD. Cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. 24 hours before the experiments, the culture media were changed into phenol red-free media containing 10% charcoal-stripped fetal bovine serum. For stimulation experiments, various steroid hormones dissolved in 100% ethanol were added into the culture media, at a final concentration of 10^{-8} M. Cells stimulated with 100% ethanol were included as controls. The cells were cultured for 24 hours, then harvested for mRNA extraction

Reverse transcriptase polymerase chain reaction

Total RNA was extracted from the breast cancer cells using Trizol reagent (Gibco BRL) following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. 2 µg of total RNA was reverse-transcribed into first strand cDNA using the Superscript™ preamplification system (Gibco BRL). The final volume was 20 µl. Based on the combined information obtained from the predicted genomic structure of the new gene and the EST sequences, two gene-specific primers were designed (Table 13) and PCR was carried out in a reaction mixture containing 1 µl of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ, USA) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 94°C for 9 minutes to activate the Taq Gold DNA polymerase, followed by 43 cycles of 94°C for 30 s, 63°C for 1 minute and a final extension at 63°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. All primers for RT-PCR spanned at least 2 exons to avoid contamination by genomic DNA.

To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The inserts were sequenced from both directions using vector-specific primers, with an automated DNA sequencer.

Structure analysis

Multiple alignment was performed using the Clustal X software package available at: ftp://ftp.ebi.ac.uk/pub/software/dos/clustalw/clustalx/ (clustalx1.64b.msw.exe) and the multiple alignment program available from the Baylor College of Medicine (BCM), Houston, TX, USA (kiwi.imgen.bcm.tmc. edu:8808/search-launcher/launcher/html). Phylogenetic studies were performed using the Phylip software package available at: http://evolution.genetics.washington.edu/phylip/getme.html. Distance matrix analysis was performed using the "Neighbor-Joining/UPGMA" program and parsimony analysis was done using the "Protpars" program. Hydrophobicity study was performed using the BCM search launcher programs (http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html). Signal peptide was predicted using the "SignalP" server (http://www.cbs.dtu.dk/services/signal). Protein structure

analysis was performed by "SAPS" (structural analysis of protein sequence) program (http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html).

RESULTS

Computer analysis of the genomic sequence predicted a putative new gene consisting of four exons. This gene was detected by all programs used and all exons had high prediction scores. EST sequence homology search of the putative exons against the human EST database (dbEST) revealed nine expressed sequence tag (EST) clones from different tissues with >95 % identity to the putative exons of the gene (Table 12). Positive clones were obtained and the inserts were sequenced from both directions. The "Blast 2 sequences" program was used to compare the EST sequences with the predicted exons, and final selection of the exon-intron splice sites was done according to the EST sequences. The presence of many areas of overlap between the various EST sequences allowed further verification of the structure of the new gene. The coding and genomic sequence of the gene has been deposited in GenBank (accession # AF135028). The 3' end of the gene was verified by the presence of poly A stretches that are not present in the genomic sequence at the end of two of the sequenced ESTs. One of the sequenced ESTs revealed the presence of an additional exon at the 5'end. The nucleotide sequence of this exon matches exactly with the genomic sequence. To further identify the 5' end of the gene, 5' RACE was performed but no additional sequence could be obtained. However, as is the case with other kallikreins, the presence of further up-stream untranslated exon(s) could not be excluded.

Mapping and chromosomal localization of the KLK-L2 gene

Alignment of KLK-L2 gene and the sequences of other known kallikrein genes within the 300 Kb area of interest enabled precise localization of all genes and determination of the direction of transcription, as shown by the arrows in Figure 13. The PSA gene was found to be the most centromeric, separated by 12,508 base pairs (bp) from KLK2, and both genes are transcribed in the same direction (centromere to telomere). The prostase/KLK-L1 gene is 26,229 bp more telomeric and transcribes in the opposite direction, followed by KLK-L2. The distance between KLK-L1 and KLK-L2 is about 35 Kilobases (Kb). The zyme gene is 5,981 bp more telomeric and the latter 3 genes are all transcribed in the same direction (Figure 13).

Structural characterization of the KLK-L2 gene and its protein product

The KLK-L2 gene, as presented in Figure 12, is formed of 5 coding exons and 4 intervening introns, spanning an area of 9,349 bp of genomic sequence on chromosome 19q13.3-q13.4. The lengths of the exons are 73, 262, 257,134, and 156 bp, respectively. The intron/exon splice sites (mGT....AGm) and their flanking sequences are closely related to the consensus splicing sites (-mGTAAGT ...CAGm-)-(32). The presumptive protein coding region of the KLK-L2 gene is formed of 879 bp nucleotide sequence encoding a deduced 293-amino acid polypeptide with a predicted molecular weight of 32 KDa. There are two potential translation initiation codons (ATG) at positions 1 and 25 of the predicted first exon (numbers refer to Figure 3). It is assumed that the first ATG will be the initiation codon, since : (1) the flanking sequence of that codon (GCGGCCATGG) matches closely with the Kozak consensus sequence for initiation of translation (GCC A/G CCATGG) (33) and is exactly the same as that of the homologous zyme gene.(2) At this initiation codon, the putative signal sequence at the N-terminus is similar to other trypsin-like serine proteases (prostase and EMSP) (Figure 14). The cDNA ends with a 328 bp of 3' untranslated region containing a conserved poly adenylation signal (AATAAA) located 11 bp up-stream of the poly A tail (at a position exactly the same as that of the zyme poly A tail)(11).

A hydrophobicity study of the KLK-L2 gene shows a hydrophobic region in the N-terminal region of the protein (Figure 15), suggesting that a presumed signal peptide is present. By computer analysis, a 29-amino acid signal peptide is predicted with a cleavage site at the carboxyl end of Ala^{29}. For better characterization of the predicted structural motif of the KLK-L2 protein, it was aligned with other members of the kallikrein multi-gene family, (Figure 14), and the predicted signal peptide cleavage site was found to match with the predicted signal cleavage sites of zyme (11), KLK1(1), and KLK2(8). Also, sequence alignment supports, by analogy, the presence of a cleavage site at the carboxyl end of Ser^{66}, which is the exact site predicted for cleavage of the activation peptide of all the other kallikreins aligned in Figure 14. Interestingly, the starting amino acid sequence of the mature protein (I I N G (S) D C) is conserved in the prostase and enamel matrix serine proteinase 1 (EMSP) genes. Thus, like other kallikreins, KLK-L2 is likely also synthesized as a preproenzyme that contains an N-terminal signal peptide (prezymogen) followed by an activation peptide and the enzymatic domain.

The presence of aspartate (D) in position 239 suggests that KLK-L2 will possess a trypsin-like cleavage pattern like most of the other kallikreins (e.g., KLK1, KLK2, TLSP,

neuropsin, zyme, prostase, and EMSP) but different from PSA which has a serine (S) residue in the corresponding position, and is known to have a chymotrypsin like activity (Figure 14). The dotted region in Figure 14 indicates an 11-amino acid loop characteristic of the classical kallikreins (PSA, KLK1, and KLK2) but not found in KLK-L2 or other members of the kallikrein-like gene family (34).

Homology with the kallikrein multi-gene family

The mature 227-amino acid sequence of the predicted protein was aligned against the GenBank database and the known kallikreins using the "BLASTP" and "BLAST 2 sequence" programs. KLK-L2 is found to have 54% amino acid sequence identity and 68% similarity with the enamel matrix serine proteinase 1 (EMSP1) gene, 50% identity with both trypsin like serine protease (TLSP) and neuropsin genes and 47%, 46%, and 42% identity with trypsinogen, zyme, and PSA genes, respectively. The multiple alignment study shows that the typical catalytic triad of serine proteases is conserved in the KLK-L2 gene (H^{108}, D^{153}, and S^{245}) and, as the case with all other kallikreins, a well conserved peptide motif is found around the amino acid residues of the catalytic triad [i.e., histidine (WLLTAA***H***C), serine(GD***S***GGP), and aspartate(***D***LMLI)] (10, 11).

Twelve cysteine residues are present in the putative mature KLK-L2 protein, ten of them are conserved in all the serine proteases that are aligned in Figure 14, and would be expected to form disulphide bridges. The other two cysteines (C^{178} and C^{279}) are not found in PSA, KLK1, KLK2 or trypsinogen, however, they are found in similar positions in prostase, EMSP1, zyme, neuropsin, and TLSP genes and are expected to form an additional disulphide bond. Twenty nine "invariant" amino acids surrounding the active site of serine proteases have been described. Of these, twenty-six are conserved in KLK-L2. One of the non-conserved amino acids (Ser^{210} instead of Pro) is also found in prostase and EMSP1 genes, the second (Leu^{103} instead of Val) is also found in TLSP gene, and the third (Val^{174} instead of Leu) is also not conserved in prostase or EMSP1 genes. According to protein evolution studies, each of these amino acid changes represents a conserved evolutionary substitution to a protein of the same group.

Evolution of the KLK-L2 gene

To predict the phylogenetic relatedness of the KLK-L2 gene with other serine proteases, the amino acid sequences of the kallikrein genes were aligned together using the "Clustal X"

multiple alignment program and a distance matrix tree was predicted using the Neighbor-joining/UPGMA method (Figure 15). Phylogenetic analysis separated the classical kallikreins (KLK1, KLK2, and PSA) and grouped the KLK-L2 with KLK-L1, EMSP1, and TLSP.

Tissue expression of the KLK-L2 gene

As shown in Table 14 and Figure 16, the KLK-L2 gene is primarily expressed in the brain, mammary gland, and testis but lower levels of expression are found in many other tissues. In order to verify the RT-PCR specificity, the PCR products were cloned and sequenced.

Hormonal regulation of the KLK-L2 gene

A steroid hormone receptor positive breast cancer cell line (BT-474) was used as a model to verify whether the KLK-L2 gene is under steroid hormone regulation. PSA was used as a control known to be upregulated by androgens and progestins and pS2 as an estrogen upregulated control. The results indicate that KLK-L2 is up-regulated by estrogens and progestins (Figure 17).

Discussion

With the aid of computer programs for gene prediction and the available EST database, a new gene, named KLK-L2 (for kallikrein like gene 2) was identified. The 3' end of the gene was verified by the presence of "poly A" stretches in the sequenced ESTs which were not found in the genomic sequence, and the start of translation was identified by the presence of a start codon in a well conserved consensus Kozak sequence.

As is the case with other kallikreins, the KLK-L2 gene is composed of 5 coding exons and 4 intervening introns and, except for the second coding exon, the exon lengths are comparable to those of other members of the kallikrein gene family (Figure 11). The exon-intron splice junctions were identified by comparing the genomic sequence with the EST sequence and were further confirmed by the conservation of the consensus splice sequence (-mGT.....AGm-) (32), and the fully conserved intron phases, as shown in Figure 11. Furthermore, the position of the catalytic triad residues in relation to the different exons is also conserved (Figure 11). As is the case with most other kallikreins, except PSA and HSCCE, KLK-L2 is more functionally related to trypsin than to chymotrypsin (34). The wide range of tissue expression of KLK-L2 should not be surprising since, by using the more sensitive RT-PCR technique instead of Northern blot analysis, many kallikrein genes were found to be expressed in a wide variety of tissues including salivary gland, kidney, pancreas, brain, and tissues of the reproductive system

(uterus, mammary gland, ovary, and testis) (34). KLK-L2 is highly expressed in the brain. Another kallikrein, neuropsin, was also found to be highly expressed in the brain and has been shown to have important roles in neural plasticity in mice (35). Also, the zyme gene is highly expressed in the brain and appears to have amyloidogenic potential (11). Taken together, these data point out to a possible role of KLK-L2 in the central nervous system.

It was initially thought that each kallikrein enzyme has one specific physiological substrate. However, the increasing number of substrates, which purified proteins can cleave *in vitro*, has led to the suggestion that they may perform a variety of functions in different tissues or physiological circumstances. Serine proteases encode protein cleaving enzymes that are involved in digestion, tissue remodeling, blood clotting etc., and many of the kallikrein genes are synthesized as precursor proteins that must be activated by cleavage of the propeptide. The predicted trypsin-like cleavage specificity of KLK-L2 makes it a candidate activator of other kallikreins or it may be involved in a "cascade" of enzymatic reactions similar to those found in fibrinolysis and blood clotting (36).

In conclusion, a new member of the human kallikrein gene family, KLK-L2 was characterized. This gene is hormonally regulated and it is mostly expressed in the brain, mammary gland and testis. KLK-L2 may be useful as a tumor marker.

Having illustrated and described the principles of the invention in a preferred embodiment, it should be appreciated to those skilled in the art that the invention can be modified in arrangement and detail without departure from such principles. All modifications coming within the scope of the following claims are claimed.

All publications, patents and patent applications referred to herein are incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

FULL CITATIONS FOR REFERENCES REFERRED TO IN THE SPECIFICATION

1. Evans BAE, Yun ZX, Close JA, Tregear GW, Kitamura N, Nakanishi S, et al. Structure and chromosomal localization of the human renal kallikrein gene. Biochemistry 1988;27:3124-3129.

2. Clements JA. The glandular kallikrein family of enzymes: Tissue-specific and hormonal regulation. Endocr Rev 1989;10:393-419.

3. Evans BA, Drinkwater CC, Richards RI. Mouse glandular kallikrein genes: Structure and partial sequence analysis of the kallikrein gene locus. J Biol Chem 1987;262:8027-8034.

4. Drinkwater CC, Evans BA, Richards RI. Kallikreins, kinins and growth factor biosynthesis. Trends Biochem Sci 1988b;13:169-172.

5. Ashley PL, MacDonald RJ. Tissue-specific expression of kallikrein-related genes in the rat. Biochemistry 1985;24:4520-5427.

6. Gerald WL, Chao J, Chao L. Sex dimorphism and hormonal regulation of rat tissue kallikrein mRNA. Biochim Biophys Acta 1986;867:16-23.

7. Riegman PHJ, Vlietstra RJ, van der Korput JAGM, Romijn JC, Trapman J. Characterization of the prostate-specific antigen gene: A novel human kallikrein-like gene. Biochem Biophys Res Commun 1989;159:95-102.

8. Schedlich LJ, Bennetts BH, Morris BJ. Primary structure of a human glandular kallikrein gene. DNA 1987;6:429-437.

9. Riegman PH, Vlietstra RJ, Suurmeijer L, Cleutjens CBJM, Trapman J. Characterization of the human kallikrein locus. Genomics 1992;14:6-11.

10. Anisowicz A, Sotiropoulou G, Stenman G, Mok SC, Sager R. A novel protease homolog differentially expressed in breast and ovarian cancer. Mol Med 1996;2:624-636.

11. Little SP, Dixon EP, Norris F, Buckley W, Becker GW, Johnson M, et al. Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. J Biol Chem 1997;272:25135-25142.

12. Yamashiro K, Tsuruoka N, Kodama S, Tsujimoto M, Yamamura Y, Tanaka T, et al. Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain. Biochim Biophys Acta 1997;1350:11-14.

13. Liu XL, Wazer DE, Watanabe K, Band V. Identification of a novel serine protease-like gene, the expression of which is down-regulated during breast cancer progression. Cancer Res 1996;56:3371-3379.

14. Luo L, Herbrick J-A, Scherer SW, Beatty B, Squire J, Diamandis EP. Structural characterization and mapping of the normal epithelial cell-specific 1 gene. Biochem Biophys Res Commun 1998;247:580-586.

15. Milanesi L, Kolchanov N, Rogozin I, Kel A, Titov I. Sequence functional inference. In: "Guide to human genome computing", ed. M.J. Bishop, Academic Press, Cambridge, 1994, 249-312.

16. Burset M, Guigo R. Evaluation of gene structure prediction programs. Genomics 1996;34:353-367.

17. Nadeau J, Grant P, Kosowsky M. Mouse and human homology map. Mouse Genome 1991;89:31-36.

18. Schachter M. Kallikreins (kininogenases) - a group of serine proteases with bioregulatory actions. Pharmacol Rev 1980;31:1-17.

19. Morris BJ, Catanzaro DF, Richards RI, Mason AJ, Shine J. Kallikrein and renin: Molecular biology and biosynthesis. Clin Sci 1981;61:351s-353s.

20. Richards RI, Catanzaro DF, Mason AJ, Morris BJ, Baxter JD, Shine J. Mouse glandular kallikrein genes. Nucleotide sequence of cloned cDNA coding for a member of the kallikrein arginyl estero-peptidase group of serine proteases. J Biol Chem 1982;257:2758-2761.

21. Van Leeuwen BH, Evans BA, Tregear GW, Richards RI. Mouse glandular kallikrein genes. Identification, structure and expression of the renal kallikrein gene. J Biol Chem 1986; 261:5529-5535.

22. Evans BA, Richards RI. Genes for the α and γ subunits of mouse nerve growth factor. EMBO J 1985; 4:133-138.

23. Rogozin IB, Milanesi L, Kolchanov NA. Gene structure prediction using information on homologous protein sequence. Comput Applic Biosci 1996;12:161-170.

24. Diamandis, E.P. Prostate specific antigen-its usefulness in clinical medicine. Trends Endocrinol. Metab., 9: 310-316, 1998.

25. Diamandis, E. P., Yu H., and Sutherland, D.J. Detection of prostate-specific antigen

immunoreactivity in breast tumours. Breast Cancer Res.Treat., 32: 301-310, 1994

26. Ishikawa, T., Kashiwagi, H., Iwakami, Y., et al. Expression of alpha-fetoprotein and prostate specific antigen genes in several tissues and detection of mRNAs in normal circulating blood by reverse trancriptase-polymerase chain reaction. Jpn. J. Oncol., 28:723-728, 1998.

27. Irwin, D.M., Robertson, K.A., and MacGillivary, R.T. Structure and evolution of the

28. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. Sequence bovine prothrombin gene. J. Mol. Biol., 212: 31-45, 1988.analysis and expression of human neuropsin cDNA and gene. Gene, 213: 9-16, 1998.

29. Goyal, J., Smith, K.M., Cowan, J.M., et al. The role of NES1 serine protease as a novel tumor suppressor. Cancer Res., 58: 4782-4786, 1998.

30. Diamandis, E.P., and Yu, H. New biological functions of prostate specific antigen? J. Clin. Endocrinol. Metab., 80 : 1515-1517, 1995.

31. Reifenberger, J., reifenberger, G., Liu, L., James, C.D. et al. Molecular genetic analysis of oligodendroglial tumors shows preferential allelic deletions on 19q and 1p.Am. J. Pathol.,145: 1175-90, 1994.

32. Iida, Y. (1990). Quantification analysis of 5'-splice signal sequence in mRNA precursors. Mutations in 5'-splice signal sequence of human β-globin gene and β-thalassemia. J. Theor. Biol. 145: 523-533.

33. Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Intimations of translational control. J. Cell Biol. 115: 887-892.

34. Clements, J. (1997). The molecular biology of the kallikreins and their roles in inflamation. *In*: S. Farmer (ed.), The kinin system, pp. 71 –97. New York: Academic Press.

35. Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998). Sequence analysis and expression of human neuropsin cDNA and gene. Gene 213 :9-16.

36. Takayama, T. K, Fujikawa, K., Davie, E. W. (1997). Characterization of the precursor of prostate-specific antigen. Activation by trypsin and by human glandular kallikrein. J. Biol. Chem. 272: 21582-21588.

Table 1. Exon or gene prediction programs used in this study[1]

No.	Program name	Source	Website or e-mail address
1	GeneBuilder (gene prediction)	Institute of Advanced Biomedical Technologies	http://l25.itba.mi.cnr.it/~webgene/genebuilder.html
2	GeneBuilder(exon prediction)	Institute of Advanced Biomedical Technologies	http://l25.itba.mi.cnr.it/~webgene/genebuilder.html
3	ORF gene	Institute of Advanced Biomedical Technologies	http://l25.itba.mi.cnr.it/~webgene/wwworfgene2.html
4	GENEID-3	BioMolecular Engineering Research Center, Boston University	http://apolo.imim.es/geneid.html (geneid@darwin.bu.edu)
5	Grail 2	Oak Ridge National Laboratory	http://compbio.ornl.gov
6	FGENEH	Baylor College of Medicine, Houston, Texas	http://mcrb.bcm.tmc.edu

1. In the final analysis of the sequences programs 1, 2, 4 and 5 only were used.

Table 2. Predicted exons of the putative gene KLK-L1. The translated protein sequences of each exon (open reading frame) are shown.

Exon No.[1]	Putative coding region[2] From(bp)	To (bp)	No. of bases	Translated protein sequence	EST match[3]	Intron phase[4]	Stop codon[5]	Catalytic triad[6]	Exon prediction program[7]
2	2263	2425	163	SLVSGSCSQIINGEDCSPHSQPWQAALVMENELFCSGV LVHPQWVLSAAHCFQ	+	II	-	H	A,B,D
3	2847	3109	263	NSYTIGLGLHSLEADQEPGSQMVEASLSVRHPEYNRPL LANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSG WGLLANGELT	+	I	-	D	A,B,C,D
4	3180	3317	137	GRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGG GQDQKDSCN	+	0	-		A,B,C,D
5	4588	4737	150	GDSGGPLICNGYLQGLVSFGKAPCGQVGVPGVYTNLC KFTEWIEKTVQAS	+	-	+	S	A,B,C

1. Conventional numbering of exons in comparison to the five coding exons of PSA,as described in Ref.14.
2. Nucleotide numbers refer to the related contig (see text and figure 1).
3. (+) =>95% homology with published human EST sequences.
4. Intron phase:0=the intron occurs between codons;I=the intron occurs after the first nucleotide of the codon; II=the intron occurs after the second nucleotide of the codon.
5. (+) denotes the exon containing the stop codon.
6. H=histidine,D=aspartic acid,S=serine.The aminoacids of the catalytic triad are bold and underlined.
7. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2, D = GENEID-3

Table 3. Predicted exons of the putative gene KLK-L2. The translated protein sequences of each exon (open reading frame) are shown*

Exon No.[1]	Putative coding sequence[2] From(bp)	To(bp)	No. of bases	Translated protein sequence	EST match[3]	Intron phase[4]	Stop codon[5]	Catalytic triad[6]	Exon prediction program[7]
1	15,361	15,433	73	MATARPPWMWVLCALITALLLGVT	+	I	-	-	-
2	17,904	18,165	262	EHVLANNDVSCDHPSNTVPSGSNQDLGAGAGEDARSDDSSSRIIN GSDCDMHTQPWQAALLLRPNQLYCGAVLVHPQWLLTAAHCRK K	+	II	-	H	A,B,C,D
3	18,903	19,159	257	VFRVRLGHYSLSPVYESGQQMFQGVKSIPHPGYSHPGHSNDLMLI KLNRRIRPTKDVRPINVSSHCPSAGTKCLVSGWGTTKSPQ	+	I	-	D	C,D
4	19,245	19,378	134	VHFPKVLQCLNISVLSQKRCEDAYPRQIDDTMFCAGDKAGRDSC Q	+	0	-	-	B,C
5	24,232	24,384	153	GDSGGPVVCNGSLQGLVSWGDYPCARPNRPGVYTNLCKFTKWI QETIQANS	+	-	+	S	A,B,C

* All footnotes same as table 2.

Table 4. Predicted exons of the putative gene KLK-L3. The translated protein sequences of each exon (open reading frame) are shown*

Exon No.[1]	Putative coding region[2] From(bp)	To(bp)	No. of bases	Translated protein sequence	EST match[3]	Intron phase[4]	Stop codon[5]	Catalytic triad[6]	Exon prediction program[7]
1	70,473	70,584	112	VHFPTPINHRGGPMEEEGDGMAYHKEALDAGCTFQDP	-	I	-	-	A,B,C,D
2	70,764	70,962	199	ACSSLTPLSLIPTPGHGWADTRAIGAEECRPNSQPWQAGLF HLTRLFCGATLISDRWLLTAAHCRK	+	II	-	H	A,B,C,D
3	73,395	73,687	293	PLTSEACPSRYLWVRLGEHHLWKWEGPEQLFRVTDFFPHP GFNKDLSANDHNDDIMLIRLPRQARLSPAVQPLNLSQTCV SPGMQCLISGWGAVSSPK	+	I	-	D	A,B,C,D
4	76,305	76,441	137	ALFPVTLQCANISILENKLCHWAYPGHISDSMLCAGLWEG GRGSCQ	+	0	-	-	A,B,C,D
5	76,884	77633	749	GDSGGPLVCNGTLAGVVSGGAEPCSRPRRPAVYTSVCHYL DWIQEIMEN	-	-	+	S	A,B

* All footnotes same as table 2.

Table 5. Predicted exons of the putative gene KLK-L4. The translated protein sequences of each exon (open reading frame) are shown*

Exon No.[1]	Putative coding region[2] From(bp)	To(bp)	No. of bases	Translated protein sequence	EST match[3]	Intron phase[4]	Stop codon[5]	Catalytic triad[6]	Exon prediction program[7]
2	24,945	25,120	176	ESSKVLNTNGTSGFLPGGYTCFPHSQPWQAALLVQGRLL CGGVLVHPKWVLTAAHCLKE	+	II	-	H	C
3	25,460	25,728	269	GLKVYLGKHALGRVEAGEQVREVVHSIPHPEYRRSPTHL NHDHDIMLLELQSPVQLTGYIQTLPLSHNNRLTPGTTCRV SGWGTTTSPQ	+	I	-	D	A,B,C,D
4	26,879	27,015	137	VNYPKTLQCANIQLRSDEECRQVYPGKITDNMLCAGTKE GGKDSCE	+	0	-	-	A,B,C,D
5	28,778	28,963	189	GDSGGPLVCNRTLYGIVSWGDFPCGQPDRPGVYTRVSRY VLWIRETIRKYETQQQKWLKGPQ	+	-	+	S	A,B,C

* All footnotes same as table 2.

Table 6. Predicted exons of the putative gene KLK-L5. The translated protein sequences of each exon (open reading frame) are shown*

Exon No.[1]	Putative coding region[2] From(bp)	To(bp)	No. of bases	Translated protein sequence	EST match[3]	Intron phase[4]	Stop codon[5]	Catalytic triad[6]	Exon prediction program[7]
2	1588	1747	160	LSQAATPKIFNGTECGRNSQPWQVGLFEGTSLRCGGV LIDHRWVL TAAHCSG	-	II	-	H	A,B,C
3	3592	3851	260	SRYWVRLGEHSLSQLDWTEQIRHSGFSVTHPGYLGAS TSHEHDLRLLRLRLRPVRVTSSVQPLPLPNDCATAGTEC HVSGWGITNHPR	+	I	-	D	A,B,C,D
4	4806	4939	134	NPFPDLLQCLNLSIVSHATCHGVYPGRITSNMVCAGG VPGQDACQ	+	0	-	-	A,B,C,D

* All footnotes same as table 2.

Table 7. Predicted exons of the unknown gene UG . The translated protein sequences of each exon (open reading frame) are shown*

Exon No.	Putative coding region[1] From(bp)	To(bp)	No. of bases	Translated protein sequence	EST match[2]	Intron phase[3]	Stop codon[4]	Exon prediction program[5]
1	44,129	44,641	513	PPLSLEPAVPERRTLRNRRSLAALAPLTPDMLLLLLPLL WGRERAEGQTSKLLTMQSSVTVQEGLCVHVPCSFSYPS HGWIYPGPVVHGYWFREGANTDQDAPVATNNPARAV WEETRDRFHLLGDPHTKNCTLSIRDARRSDAGRYFFRM EKGSIKWNYKHH RLSVNVT	+	I	-	B,C
2	44,843	45,121	279	ALTHRPNILIPGTLESGCPQNLTCSVPWACEQGTPPMIS WIGTSVSPLDPSTTRSSVLTLIPQPQDHGTSLTCQVTFPG ASVTTNKTVHLNVS	+	I	-	A,B,C,D
3	45,327	45,374	48	YPPQNLTMTVFQGDGT	-	I	-	A,B,D
4	46,318	46,542	225	EGQSLRLVCAVDAVDSNPPARLSLSWRGLTLCPSQPSN PGVLELPWVHLRDAAEFTCRAQNPLGSQQVYLNVSLQ	+	I	-	A,B,C
5	47,195	47,283	186	SKATSGVTQGVVGGAGATALVFLSFCVIFV	+	0	-	A,B,C,D
6	49,136	49,554	186	GPLTEPWAEDSPPDQPPPASARSSVGEGELQYASLSFQ MVKPWDS RGQEATDTEYSEIKIHR	+	-	+	A,B,C

* All footnotes same as table 2.

Table 8 . Homology between the predicted amino acid sequences of the newly identified putative genes and protein sequences deposited in Genbank

No.	Gene identity	Homolgous known protein	Identity% (number of amino acids)
1	KLK-L1	• Human stratum corneum chymotryptic enzyme	44(101/227)
		• Rat kallikrein	40(96/237)
		• Mouse glandular kallikrein K22	39(94/236)
		• Human glandular kallikrein	38(93/241)
		• Human prostatic specific antigen	37(91/241)
		• Human protease M	37(87/229)
2	KLK-L2	• Human neuropsin	48(106/219)
		• Human stratum corneum chymotryptic enzyme	47(103/216)
		• Human protease M	45(99/219)
		• Human trypsinogen I	45(100/221)
		• Rat trypsinogen	44(98/220)
3	KLK-L3	• Human neuropsin	44(109/244)
		• Rat trypsinogen 4	39(95/241)
		• Human protease M	38(98/253)
		• Human glandular kallikrein	37(94/248)
		• Human prostatic specific antigen	36(89/242)
4	KLK-L4	• Human protease M	52(118/225)
		• Human neuropsin	51(116/225)
		• Mouse neuropsin	51(116/226)
		• Human glandular kallikrein	48(113/234)
		• Human prostatic specific antigen	47(108/227)
5	KLK-L5	• Human neuropsin	44(81/184)
		• Rat trypsinogen I	42(76/178)
		• Rat trypsinogen II	42(75/178)
		• Human protease M	41(73/178)
6	UG	• Human myeloid cell surface antigen CD33	61(144/233)
		• Human OB binding protein-2	50(166/328)
		• Human OB binding protein-1	43(189/431)
		• Human myelin associated glycoprotein	27(86/311)

Table 9. Expressed sequence tags with >95% homology to exons of the prostase/KLK-L1 gene.

2

GenBank #	Source	Tissue	homologous exons
AA551449	I.M.A.G.E.	prostate	3,4,5
AA533140	I.M.A.G.E.	prostate	4,5
AA503963	I.M.A.G.E.	prostate	5
AA569484	I.M.A.G.E.	prostate	5
AA336074	TIGR	endometrium	2,3

5

Table 10. Primers used for reverse transcription-polymerase chain reaction (RT-PCR) analysis of various genes.

Gene	Primer name	Sequence[1]	Product size (base pairs)
Prostase (KLK-L1)	RS	TGACCCGCTGTACCACCCCA	278
	RAS	GAATTCCTTCCGCAGGATGT	
pS2	PS2S	GGTGATCTGCGCCCTGGTCCT	328
	PS2AS	AGGTGTCCGGTGGAGGTGGCA	
PSA	PSAS	TGCGCAAGTTCACCCTCA	754
	PSAAS	CCCTCTCCTTACTTCATCC	
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT	372
	ACTINAS	TCTCCTTAATGTCACGCACGA	

1. All nucleotide sequences are given in the 5'→3' orientation.

Table 11. Tissue expression of prostase/KLK-L1 by RT-PCR analysis

Expression level			
High	**medium**	**low**	**No Expression**
Prostate	Mammary gland	Salivary glands	Stomach
Testis	Colon	Lung	Heart
Adrenals	Spinal cord	Brain	Spleen
Uterus		Bone marrow	Placenta
Thyroid		Thymus	Liver
		Trachea	Pancreas
		Cerebellum	Kidney
			Fetal brain
			Fetal liver
			Skeletal muscle
			Small intestine

Table 12. EST clones with >95% homology to exons of KLK-L2

GENBANK # ; Tissue of Origin		I.M.A.GE. ID	Homologous exons
W73140	Fetal heart	344588	4,5
W73168	Fetal heart	344588	3,4,5
AA862032	Squamous cell carcinoma	1485736	4,5
AI002163	Testis	1619481	3,4,5
N80762	Fetal lung	300611	5
W68361	Fetal heart	342591	5
W68496	Fetal heart	342591	5
AA292366	Ovarian tumor	725905	1,2
AA394040	Ovarian tumor	726001	5

Table 13. Primers used for reverse transcription polymerase chain reaction (RT-PCR) analysis.

Gene	Primer name	Sequence[1]	Product size (base pairs)
KLK-L2	KS	GGATGCTTACCCGAGACAGA	342
	KAS	GCTGGAGAGATGAACATTCT	
pS2	PS2S	GGTGATCTGCGCCCTGGTCCT	328
	PS2AS	AGGTGTCCGGTGGAGGTGGCA	
PSA	PSAS	TGCGCAAGTTCACCCTCA	754
	PSAAS	CCCTCTCCTTACTTCATCC	
Actin	ACTINS	ACAATGAGCTGCGTGTGGCT	372
	ACTINAS	TCTCCTTAATGTCACGCACGA	
KLK-L2	R1	CCGAGACGGACTCTGAAAACTTTCTTCC	
	R2	TGAAAACTTTCTTCCTGCAGTGGGCGGC	

1. All nucleotide sequence are given in the 5'→3' orientation.

Table 14. Tissue expression of KLK-L2 by RT-PCR analysis.

Expression level			
high	**medium**	**low**	**No Expression**
Brain	Salivary gland	Uterus	Stomach
Mammary gland	Fetal brain	Lung	Adrenal gland
Testis	Thymus	Heart	Colon
	Prostate	Fetal liver	Skeletal muscle
	Thyroid	Spleen	
	Trachea	Placenta	
	Cerebellum	Liver	
	Spinal cord	Pancreas	
		Small intestine	
		Kidney	
		Bone marrow	

We Claim:

1. An isolated nucleic acid molecule which comprises:

(i) a nucleic acid sequence encoding a protein having substantial sequence identity preferably at least 60% sequence identity, with an amino acid sequence of KLK-L1-KLK-L6 as shown in Tables 2 to 6 or Figure 18;

(ii) a nucleic acid sequence encoding a protein comprising with an amino acid sequence of KLK-L1-KLK-L6 as shown in Tables 2 to 6 or Figure 18;

(iii) nucleic acid sequences complementary to (i);

(iv) a degenerate form of a nucleic acid sequence of (i);

(v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);

(vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising with an amino acid sequence of KLK-L1-KLK-L6 as shown in Tables 2 to 6 or Figure 18; or

(vii) a fragment, or allelic or species variation of (i), (ii) or (iii).

ABSTRACT OF THE DISCLOSURE

The invention relates to nucleic acid molecules, proteins encoded by such nucleic acid molecules; and use of the proteins and nucleic acid molecules

5

VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS (37 CFR 1.9(f) AND 1.27 (d)) - NONPROFIT ORGANIZATION	Docket No.

Serial No.	Filing Date	Patent No.	Issue Date

Applicant/ Patentee: **George M. Yousef and Eleftherios P. Diamandis**

Invention: **Novel Human Kallikrein-Like Genes**

I hereby declare that I am an official empowered to act on behalf of the nonprofit organization identified below:

NAME OF ORGANIZATION: **Mount Sinai Hospital**

ADDRESS OF ORGANIZATION: **600 University Avenue**
Toronto, Ontario
Canada
M5G 1X5

TYPE OF NONPROFIT ORGANIZATION:

- ☐ University or other Institute of Higher Education
- ☐ Tax Exempt under Internal Revenue Service Code (26 U.S.C. 501(a) and 501(c)(3))
- ☐ Nonprofit Scientific or Educational under Statute of State of The United States of America
  Name of State: Citation of Statute:
- ☒ Would Qualify as Tax Exempt under Internal Revenue Service Code (26 U.S.C. 501(a) and 501(c)(3)) if Located in The United States of America
- ☐ Would Qualify as Nonprofit Scientific or Educational under Statute of State of The United States of America if Located in The United States of America
  Name of State: Citation of Statute:

I hereby declare that the above-identified nonprofit organization qualifies as a nonprofit organization as defined in 37 C.F.R. 1.9(e) for purposes of paying reduced fees to the United States Patent and Trademark Office regarding the invention described in:

- ☒ the specification to be filed herewith.
- ☐ the application identified above.
- ☐ the patent identified above.

I hereby declare that rights under contract or law have been conveyed to and remain with the nonprofit organization with regard to the above identified invention.

If the rights held by the above-identified nonprofit organization are not exclusive, each individual, concern or organization having rights to the invention is listed on the next page and no rights to the invention are held by any person, other than the inventor, who could not qualify as an independent inventor under 37 CFR 1.9(c) or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern or organization to which I have assigned, granted, conveyed, or licensed or am under an obligation under contract or law to assign, grant, convey, or license any rights in the invention is listed below:

☒ no such person, concern or organization exists.
☐ each such person, concern or organization is listed below.

FULL NAME _____
ADDRESS _____
☐ Individual ☒ Small Business Concern ☐ Nonprofit Organization

FULL NAME _____
ADDRESS _____
☐ Individual ☐ Small Business Concern ☐ Nonprofit Organization

FULL NAME _____
ADDRESS _____
☐ Individual ☐ Small Business Concern ☐ Nonprofit Organization

FULL NAME _____
ADDRESS _____
☐ Individual ☐ Small Business Concern ☐ Nonprofit Organization

Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR 1.27)

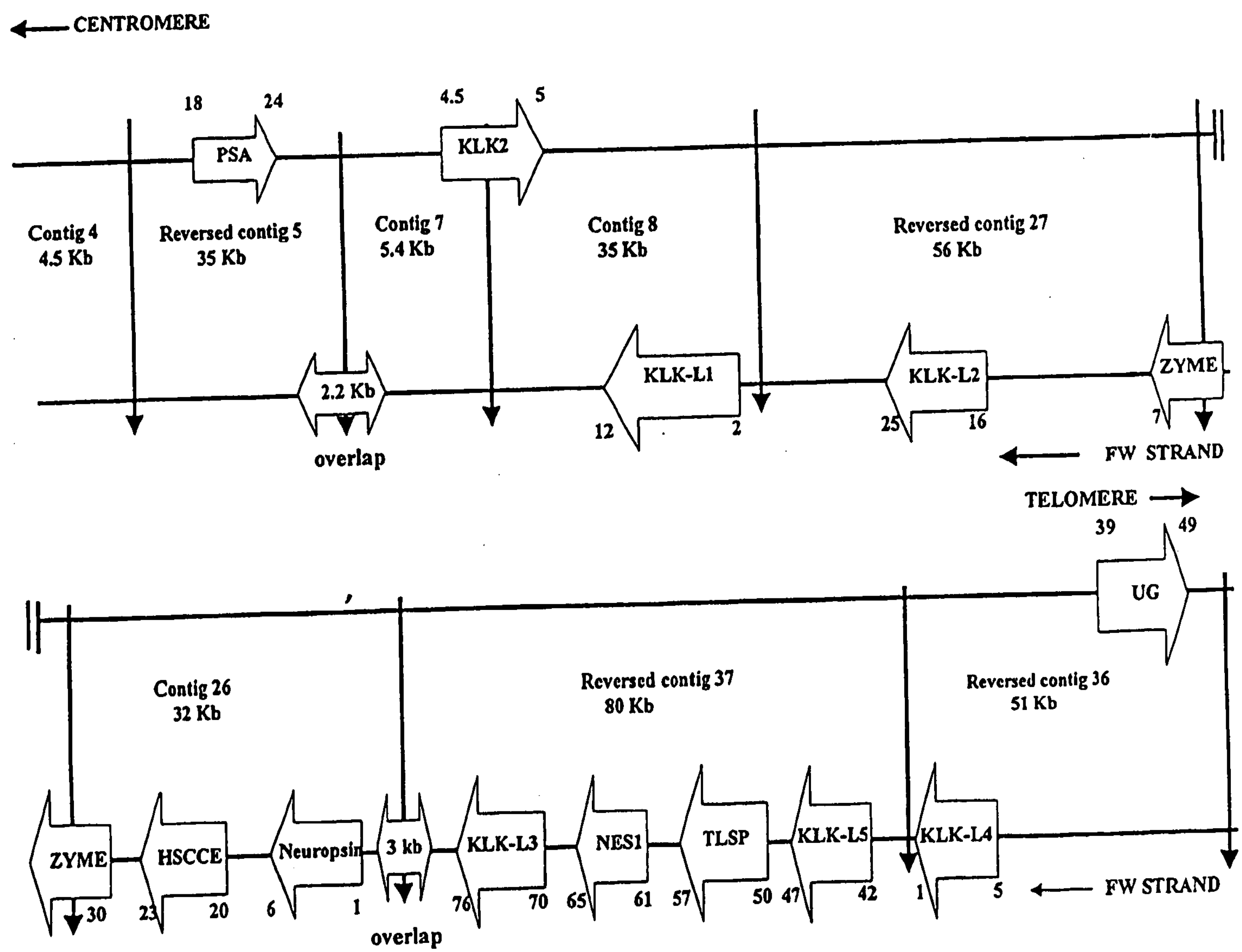
I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING: Joseph Mapa
TITLE IN ORGANIZATION: Executive Vice-President & C.O.O.
ADDRESS OF PERSON SIGNING: **Mount Sinai Hospital**
600 University Avenue
Toronto, Ontario
Canada M5G 1X5

SIGNATURE: _____ DATE: July 20, 1999

FIGURE 1



CENTROMERE
18
24
PSA
4.5
5
KLK2
Contig 4
4.5 Kb
Reversed contig 5
35 Kb
Contig 7
5.4 Kb
Contig 8
35 Kb
Reversed contig 27
56 Kb
2.2 Kb
overlap
KLK-L1
12
2
KLK-L2
25
16
ZYME
7
FW STRAND
TELOMERE
39
49
UG
Contig 26
32 Kb
Reversed contig 37
80 Kb
Reversed contig 36
51 Kb
ZYME
30
HSCCE
23
20
Neuropsin
6
1
3 kb
overlap
KLK-L3
76
70
NES1
65
61
TLSP
57
50
KLK-L5
47
42
KLK-L4
1
5
FW STRAND

FIGURE 2

KLK – L 1

TATCTCATGAGAGAGAATAAGAACATGAAAAGAGAAAGAATGAGAGAGAG
AGAGAGAAAGAAAAAGGAGAGTGGAGTCTAGGATCTGGGCAGGGGTCTCC
TCCCTGGGTCCCTAGACCCTGCTGCCAGCCCCTTCTGGGCCCCCAACCAC
TGCCTGGTCAGAGTTGAGGCAGCCTGAGAGAGTTGAGCTGGAAGTTTGCA
GCACCTGACCCCTGGAACACATCCCCTGGGGGCAGGCCAGCCCAGGCTGA
GGATGCTTATAAGCCCCAAGGAGGCCCCTGCGGAGGCAGCAGGCTGGAGC
TCAGCCCAGCAGTGGAATCCAGGAGCCCAGAGGTGGCCGGGTAAGAGGCC
TGGTGGTCCCCCACTAAAAGCCTGCAGTGTTCATGATCCAACTCTCCCTA
CAGCTCCATGTCGCTGGATTCTCAGCCTCTGTGCCTTCTGTCTCCACATC
TCTCTAGACAGATCTCTCACTGTCTCTAGTTAGGAGTCACTGTCTCTAGT
TAGGGGTCTCTCTGTCTCTCTGAATCTATATCTCCATGTCTAACTCTCAG
ACTGTCTCTGAGGATATCTCTCAAGCACTCTGTCTCTCCGGCTCTGATTC
TCTGTGTGTCTTCCCTCCATGCTTGTTTGTGGGTGGCTAGACACCATCTC
TCCCCATTCACAGATGGCTAGATGCTTTCTCTAAACTTTCCTTTCTACCT
AGTTCTCTCTCTCTCTCTTTTCCCATCTCTCTCTCTCTTTTTCTCTCTCA
GTCTCTAAATCTGTCTCTCTAGGTTCTGGGTCCATGGATGGGAGAGGGGG
TAGATGGTCTAGGCTCTTGCCTACCTAATAACGTCCCAGAGGGAAGAAAG
GGAGGGACAAAGAGAGGGATGGAGAGACTTGGGCTGAAGATCCCCAGACA
CGGCTAAGTCTCAGTCCTCATCCCCAGGTGCTGACGTGATGGCCACAGCA
GGAAATCCCTGGGGCTGGTTCCTGGGGTACCTCATCCTTGGTGTCGCAGG
TATCTGAGTATGCGTGTGTGTGTCTGTCCGTGCTTGGGGGCACAGTGTTT
GTTAATGTTCAGGTGTGACTCAGTGTCCTCTTGCTTGTGACTGCAAAGCT
GCCTGTGAGACGGTACCGTGTTATCCGTCCGCCATGGCTGTGCCCCTGCA
ACTCCTTGTATCGTGGTAAATTTGTGTGTGGCAGTGTGCCTGGGTGTGTG
GTTGTACCTGTGAGACTCTGACAGTTTGTGCCTCTGAATATCTGGTGGAG
TGACAACAGTGTAATGATGATATGGGGACAGGGGAAGCCGAGGGTGCAGG
AGATTGTGCTTCCTGGGGCGTGATCCATTGCTGGGAATCTGTGCCTGCTT
CCTGGGTCTTCAGTCCTGAGATCCCCCTCTCCCATCCCCAAGGAACTCAC
CTCACAGGACTATAAAACGGTGTTTTGGTGTGCATGGGCTTGTGGCTTGG
TGTGACTGTGGGCAAGGCTGGGAGAGGATAGGAGTGACTCGGCGCAGGAC
CGACTCTTTGAGCATCAGTCTGCGCAGACAAGTGACCCGATCCTTGCTCC
CAGCAACAACTCCACCCCCTGAGCTTTAATTCACCCCGAAGGACCCGATC
CTACCGCTATGAGCCTAGACTCCTCTGTTGAACCCCTCCTGACCGTGGCT
TTGCACCGCGATGGCACCAGTCTCACCTCCAGAGCTCACCCCAGAGCCCT
GACTCCGCCCCAGAAGCCCTGGTCCCACCTTCTGAGACTGCCTCTAGCCA
TAACCCAGCTCTTGAAGCCTTGATGGCGCCCCTGCGCTGTAACCCCAACC
CTAGGAGCACTGATCCCGCCTTCTCAGCCCACCCCCATGCCCTGAGTCTC

FIGURE 2 (cont'd)

CTCCCAGGAGCCCTGACTACCCTGAATCCCTGACCAGGCTCCTGCACCGT
GATCACCGCCCCTGGGAGCCCTAGGCCTATATCCTGGACCAGCCCCTGAA
GCTCCGATCATGACCCCTGCACCATAACCCCACCCCCAGGAGCCCTGGGT
CCGCCCCCTGGGCCCGCCCCCAGCCCTGACTCGGCCCCCCAAGAGTCCTG
ACTGCTCCTGAAGCCCTGACCACGCCCCTGCTCGGTAACCCCTCCCCCAA
GAGCCCTGGGCCCGCCTCCTGAGCCCGTTCCCAGCCCTGACTCCGCCCCG
AGGAGCCCTGACTGCTCCTGAACCTCTGACCACGCCCCTGCTCGGTAAGC
CCACCCCCAGGAACCCTGGGCCCGCCTCCTGGTCCCGATCCCATCCCTGA
CTCCGCCCTCAG**GATCTCTCGTCTCTGGTAGCTGCAGCCAAATCATAAAC**
GGCGAGGACTGCAGCCCGCACTCGCAGCCCTGGCAGGCGGCACTGGTCAT (1)
GGAAAACGAATTGTTCTGCTCGGGCGTCCTGGTGCATCCGCAGTGGGTGC
TGTCAGCCGCACACTGTTTCCAGAAGTGAGTGCAGAGGTAGGGGGAGTGG
GCAGGGCCTGGGTCCGGGGGCGGGGCCTAATATCAGGCTCATCTTGGGGT
GCTCAGGGGGAAACAGCGGTGAAGGCTCTGGGAGGAGGACGGAATGAGCC
TGGATCCGGGGAGCCCAGAGGGAAGGGCTGGGAGGCGGGAATCTTGCTTC
GGAAGGACTCAGAGAGCCCTGACTTGAAATCTCAGCCCAGTGCTGAGTCT
CTAGTGAACTAAGGCAAGTTCTTGTCCCTGAATTTTTGTGAATGAGGATT
TGAGACCATGGTTAAGTAGCTCTTAGGGTGTTTAGCGAAGAGGGTGGGGT
TGGGGTTAGGAGATGGGGATGGGAATGGGGTTGAAGATGAGAATGGAGGT
AAGGATGTAGTTGCCACAAAACTGACCTGCCCTCCGTGGCCCACAG**CTCC**
TACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGG (2)
GAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACA
GACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTG
TCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTAC
CGCGGGGAACTCTTGCCTCGTTTCTGGCTGGGGTCTGCTGGCGAACGGTG
AGCTCACGGGTGTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCG
CGGGGGCTGACCCAGAGCTCTGCGTCCCAG**GCAGAATGCCTACCGTGCTG**
CAGTGCGTGAACGTGTCGGTGGTGTCTGAGGAGGTCTGCAGTAAGCTCTA (3)
TGACCCGCTGTACCACCCCAGCATGTTCTGCGCCGGCGGAGGGCAAGACC
AGAAGGACTCCTGCAACGTGAGAGAGGGGAAAGGGGAGGGCAGGCGACTC
AGGGAAGGGTGGAGAAGGGGGAGACAGAGACACACAGGGCCGCATGGCGA
GATGCAGAGATGGAGAGACACACAGGGAGACAGTGACAACTAGAGAGAGA
AACTGAGAGAAACAGAGAAATAAACACAGGAATAAAGAGAAGCAAAGGAA
GAGAGAAACAGAAACAGACATGGGGAGGCAGAAACACACACACATAGAAA
TGCAGTTGACCTTCCAACAGCATGGGGCCTGAGGGCGGTGACCTCCACCC
AATAGAAAATCCTCTTATAACTTTTGACTCCCCAAAAACCTGACTAGAAA
TAGCCTACTGTTGACGGGGAGCCTTACCAATAACATAAATAGTCGATTTA
TGCATACGTTTTATGCATTCATGATATACCTTTGTTGGAATTTTTTGATA
TTTCTAAGCTACACAGTTCGTCTGTGAATTTTTTTAAATTGTTGCAACTC
TCCTAAAATTTTTCTGATGTGTTTATTGAAAAAATCCAAGTATAAGTGGA
CTTGTGCAGTTCAAACCAGGGTTGTTCAAGGGTCAACTGTGTACCCAGAG
GGAAACAGTGACACAGATTCATAGAGGTGAAACACGAAGAGAAACAGGAA
AAATCAAGACTCTACAAAGAGGCTGGGCAGGGTGGCTCATGCCTGTAATC
CCAGCACTTTGGGAGGCGAGGCAGGCAGATCACTTGAGGTAAGGAGTTCA
AGACCAGCCTGGCCAAAATGGTGAAATCCTGTCTGTACTAAAAATACAAA
AGTTAGCTGGATATGGTGGCAGGCGCCTGTAATCCCAGCTACTTGGGAGG

<u>FIGURE 2 (cont'd)</u>

CTGAGGCAGGAGAATTGGTTGAATATGGAGGCAGAGGTTGAAGTGAGTT
GAGATCACACCACTATACTCCAGCTGGGGCAAGAGAGTAAGAGTCTGTCT
CAAAAAAAAAAAAAAAAAAGACTTTACAAAGAGATGCAGAGACACTGAGA
CAGATAAACAAGCCACAAAGGAGACAAAGGAGAGACAGACAAACAGAAAC
AGACAGACCACAAGCCCAAGAGAAGCAGCCAGCATTCAGGACATAGGACA
TCGGGAAGCAGGATTAGATGAAGTCAGGGATCTGGAATGGGACTTCCAAC
AGATATGTTGCTGGGCTATGTTGTTATTGATGATGGTTCTGTCTTTGTTT
CTCAGTCTCATTTAGTTCCTTTCTGAGCCCATATCCATTTCCACCTCTCT
GTGTTTTGAATTCTGACTCTCCCTCTCTTCACAACAG<u>GGTGACTCTGGGG</u>
<u>GGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTGTCTTTCGGAAAA</u> (4)
<u>GCCCCGTGTGGCCAAGTTGGCGTGCCAGGTGTCTACACCAACCTCTGCAA</u>
<u>ATTCACTGAGTGGATAGAGAAAACCGTCCAGGCCAGT</u>*TAA* STOP

FIGURE 3

KLK-L 2

GGGCCCAGAG TGAAGGCAAG AGAAGGAGTT GAGAGCTCCC TCTGCAAAGT GGCTTGAGTC
TCCCCTGCCT AAAATGCAGG GAGAGGGAGG CAGAAAGACA GGGAAGAGGA AGGGGTGGGG
AAGAAAGAGA GAGAGAGAGA GAGACAGAAT AACACAACTA CAGAAACACA GAGAGAACAC
ACAGAGAGCC TGGGACACAG GGACACACAG AGTCAGAGAG AAAAGAGAAG ATAGAGAAAG
ACACAAATGG AGACACAGAG GTGTAAAGAA AGAGAGATTA ACAGAGTCCC AGATACACGC
AAAGGGGCAG AAGCACAGTT TTCAGGGTGG TGTCTATGAT CATCTTCTTT TTTTTTTTTT
TTTTTTTTTT TTTTTGAGAC GGAGTCTCGC TCTGTCGCCC AGGCTGGAGT GCAGTGGCGG
GATCTCGGCT CACTGCAAGC TCCGCCTCCC GGGTTCACGC CATTCTCCTG CCTCAGCCTC
CCAAGTAGCT GGGACTACAG GCGCCCGCCA CTACGCCCGG CTAATTTTTT TGTATTTTTA
GTAGAGACGG GGTTTCACCG TTTTAGCCGG GATGGCCTCG ATCTCCTGAC CTCGTGATCC
GCCCGCCTCG GCCTCCCAAA GTGCTGGGAT TACAGGCGTG AGCCACCGCG CCCGGCCATG
ATCATCTTCT TGACTATGCT GATGTGACAA GTACCTAAAG CCATCAGACT CTACCCTTTA
AATATGCAGT TTGGGCCAGG CACCGTGGCT CATGCCTGTA ATTCCAGCAC TTTGGGAGGC
AGAGGTGGGT GAATCACTTG AGGCCAGGAG TTTGAGACCA GCCTGGCCAA CATGGTGAAA
CTCTGTCTTT ACTAAAAAAA AAAAAAAAAA AAAAAAAATC AGCCGGGTGT CGTGGGGCAC
ACCTGTAATC CCAGCTATGC TGGAGGCTGA GGCACGAGAG TCACTTGAAC CCTGGAGGCG
GAGGTTGCAG TGGGCCGAGA TCACATCACC GCCCTCCAGC CTGGGCGACA GAGCAAGACT
CTGTCTCAAA TAAATAAATA AACAAACGAA CAAGCAGTTT GTTGTACCTT AGTTATATCT
AAAAAAAAAA TGCTGTCAAC AAATAGAGCA GAAGTGAAAT AAAGGAAAAT AAATGGGCCA
AGAACTCTAA GGTATATTTG ACAAATCATT CAGAACCTTT AAAAAAGAAA GAATCACAGA
GGCATAGAAA GACAGGGAGG AACAGGGAGA CAGAAACACC TGTGGCCCAA GGAGAACAAA
ACAAGGCTCC TAAGACAGAC AGGAGGAGAG AGAGAGAGAG TGAGTGAGAG ACAGACAGAG
AAAAAGACAG AGAGAGAGAG ACAGAGACAG AGAGACAGAG AGGCGAGAGG GATAGAAAGA
GAGAGAGGGG TGGAGAGAGA CACGAGATAT TGAGAGAGAC TCAGAAAGAT AGCCGAGGGA
GAACCACAGA GAGATGGAAG AAGACTCTGA GAAAAAACCA GAGACAAAGA TGGAAAGAGG
AGTATCGAGG GTGAACAGAC AGTGGTGGAA TGAGCAAAAT GCAGAGAAGA AAGCAAGCAA
TCCAGGCGCC AAGAATAGTG ACCCAGAGTT GGTGAGAAGC CAGATCCTTA AGGCTGGGGG
AGGCAGGGAA GGGGCTGGCC TGGCTTCCGG AGACCCCTCC CCATTCTCCG GGCCAGGGAG
GTAGGGAGTG ACATTCCGGA CTGGGTGGGG GGTGCTCTGG GGGTGGAGAT AGGGGGAGCA
GGAGGAGCTA TTGCTAAGGC CCGATAGGCA CCTCATTGCC CGGGAATGTG CCCCAGGGAG
CAGTGGGTGG TTATAACTCA GGCCCGGTGC CCAGAGCCCA GGAGGAGGCA GTGGCCAGGA
AGGCACAGGC CTGAGAAGTC TGCGGCTGAG CTGGGAGCAA ATCCCCCACC CCCTACCTGG
GGGACAGGGC AAGTGAGACC TGGTGAGGGT GGCTCAGCAG GCAGGGAAGG AGAGGTGTCT
GTGCGTCCTG CACCCACATC TTTCTCTGTC CCCTCCTTGC CCTGTCTGGA GGCTGCTAGA
CTCCTATCTT CTGAATTCTA TAGTGCCTGG GTCTCAGCGC AGTGCCGATG GTGGCCCGTC
CTTGTGGTTC CTCTCTACCT GGGGAAATAA GGTAGGGGAG GGAGGGGAAG TGGGTTAAGG
GCTCCCCGGA TCGCCTGGGC CTCCCAACCC TCTGACATTC CCCATCCAGG TGCAGCGGCC
ATGGCTACAG CAAGACCCCC CTGGATGTGG GTGCTCTGTG CTCTGATCAC AGCCTTGCTT
CTGGGGGTCA CAGGTAACCA GAACTCTGGG GTGGGAGGGT TGTGGGATTG GGAGGACTGT
CTCTGCGGCA CTAGAGCGCC TGTCCCCTGG GGAACTGTGT GAGCCTGGGC ATGACTCCGG
GACCGGGTGA ATGTGAGTCT CTGTCTGTAC TTGTGGTTGT GCGATCGTAT GTGGCCCTGT
GACTGCCACG GTGTGTGTCG GGGAGGGGGA TGCCTTTTCC CATATCAGGT GACTGTGCGG
CAGGTGGCAC TGACCCTTTG AGGCTGTGTG TGTGGTTTTG TGATTGTGTG TGCATTTAAG
ATTGTGTGTG GCTCCACAGC TGTGTGGGTG AATGCATGTA GCACTGGGGG TGTTCACTGT
GTGTTTGGCT GTGTGTGGTG ACTTGGCATT GTATATGACT GCAGGTATCT GCAGTTCCTG
TCCCTGAGGT CCCGGGATTG CGTGCAACAA AAGTGGTCAT CACCATGGAA AGCTGTGACT
GTGTGCTGCT TGCAGGCGAT TATGTGATTG TGGCTGAGTG TGACGTTATG GATGCCCGTA
TTTGTGACCG TGTGACTACC TGAAGCTCTG TGTAGGGGTG ACTGTATGTG ACTGTGTGTG
TCTGTGTGAG GCCGTGTAAA TGCTACTGTA TGTGTGATGG TGCAGCTGTG TGTCTGGAGT
TTCTGTCTCT GCCTGGAGGG ATAGAGGGTG CAGGGGTAGC TATCTCTGGG AGATGGGTGC
CAGGTGACTG ACTTGCAGTG TGTGCCTGTG TGCAGAAGAG TATGTGGCAG TCTGAACATC
TGTGCACACA CGGCATCTGT GCGTGGCACT GAGACACTGT GGATGAGGGT GTGCGATCCC
GCTAGGCTGC CCGGGAGCGT GTGTACCTGG AGACAGAGCT GTATGTTAGC TGCACCTGTG
GAGGCAACAT GGGCGTGTCT GCAGAACTGC GTGCGTGCTT GGCTGTTACT GCTGTTGTGC

FIGURE 3 (cont'd)

GCGTGGTTCT TGGGGTGAGT TCGTGAATGA TGGTGGTGCC AGGGCCATCA GCAAGGGTAA
GAACCAGGCC GGGCGCGGTG GCTCACGCCT GTAATCCCAG CCCTTTGGGA GGCCGAGGCA
GGCGGATCAC CTGAGGTCGG GAGATCGAGG CCAGCCTGAC CAACATGGAG AACCCCGTCT
CTACTAAAAA TACAAAAAAT TAGCTGGTGT GGTGGCGCGT GCCTGTAATC CCAGCTACTC
GGGAGACTGG GGCAGAAAAA TCGCTTGAAC CCGGGAGGTG GAGGTTGCGG TGAGCCGAGA
TCGCGCCATT GCACTCCAGC CTGGGCAACA AGAGCGAAAC TCCGTCTCGA AAGAAAAAAA
GAAAAAAAAA AGGGTAAGAA CCAGTGAATG GGCACGGGAG GACTGATGAT GGAGTGGGGC
ATGCATGTAG TCTGTAGGTC TGTGTGTGAG AGGAGGAGAT TGACAGGATT GAGAAGGCAT
GTTTTCATCT GAGAATTCAG AAACCTAGGC CTGCTCTTCC CCTCCATGTG GCCCCCTAAG
CTGAGCCCTT CTTTCCTGGT CCTGCTTTCG GAACCCTAGC TCCGCCCATG AGCTCTGACC
CCACCTCCTT TCCTCAACCA CGCCCCTAGG CCAGACTCTA GTGGACCCCG CCTAAGGCCA
CACCCCTTTG GGCCAGGCTC CACCCCCTAT TCTGTGGGTA CCTTCTAGAA CCCCCTTCAA
AGTCAGAGCT TTTTTTTTTT TTTTTTTGGA GACAGTCTTG CTCTCTCTCC CAGGCTGGAG
TGCAGTGGCG TGATCTCGGC TCACTGCAAC CTCTGCCTCC CAGGTTCAAG TGATTCTCGT
GCCTCCACCT CCTGAGTAGC TGGGATTACA GGTGCGCGCC ACCACGCCTG GCTAATTTTT
GTGTCTTTAG TAGAGACAGG GTTTCACCTT GTTGGCCAGG CTGGTCTCAA ACTCCCAACC
TCAGGTGATC CGCCCACCTC GGCCTCCCAG AGTGCTGGGG TTACAGGCGT GAGCCACCGC
CCCCAGCCCA AAGTCAGAGC TCTTTATAGG AGACTCTAAC ATGTAACCCT GACCCTGGCC
CTAACTAAGT CAATTCCAAA CCCCTTCCTG CCTCCAGCCC TGACCCCACT CACTGAGGCC
TGACCCCACT TCTTGAGACC AGTTCCATCC CTAAAGCCCT GGTCTCCCTC CCATCCCCAG
GCTCCAGCCC CCACAGCTTT GGCACTACCC CTGAGCTTGT CCAGGAATCC TGTACCCAAT
TTTACCCTCA CATGTAGTTC TAGCCAATTC CAGGAATCTG TGAGGTCCAG TTAGAGTCCA
GTAACCCTAC CTGAGCCTGG GCTCTGTCCT TGAGCTTGAG CCTGGGCTTG AGAGGTGCCA
CTCTTATTGT CCAGGCCCTG CCCCTGCCCC CTCAGCATGT CAGACACCCA CCCTCTAGCT
GGTCTGGCCT CTTGAGTCTG AAACCCACCC CCAGCCCAAG CCCCGCCTCT GAGCCCCGCC
CAACCCATTT TCCGTTCCCA GAGCATGTTC TCGCCAACAA TGATGTTTCC TGTGACCACC
CCTCTAACAC CGTGCCCTCT GGGAGCAACC AGGACCTGGG AGCTGGGGCC GGGGAAGACG
CCCGGTCGGA TGACAGCAGC AGCCGCATCA TCAATGGATC CGACTGCGAT ATGCACACCC
AGCCGTGGCA GGCCGCGCTG TTGCTAAGGC CCAACCAGCT CTACTGCGGG GCGGTGTTGG
TGCATCCACA GTGGCTGCTC ACGGCCGCCC ACTGCAGGAA GAAGTGAGTG GGAGTTCCAA
GAGGAGGGTT GGTGGGGACG GGGAAGTGGG GGTGGGGGTG GGGAAGTGGG GGTGGGGGTG
TCATGGAGGT GAGGGCTGGT GGGGACGGGG AAGTGGGGTT GGGGGTGTCA TGGAAGGTGA
GGGTTGGTGG GGATGGGTTG GGGATGTGGG AGCAGGAGGA GGTCGAGTTG GGGATAGGAC
TAAGGATGGA GTTTTGCGGG GGAGCAAGGT GGGAGGATGA GGTTGGAGAG GGGAGAGTGT
TGTGGTAGGG AATGGGAAGG AGCCAAGGAT GGGTTGGATT TGGGGTTAGG AGCATATATT
TGTTGAATGG TTTGGGATGG AGGTGGAATT GGGATTGGCT TTAGAATTGG GGGTGGGTGA
AAATCGGGCT GGGGTGGAAA TGAAGATAGC ATGGAGATAG GGTTGAGATT GGGAGCAGAT
ATAGAATGAA GGATGGGGAT TGGAGTTTTG GGTGGGGTTG GAGATGGTTG GATTTGGGCT
TGAGAATGCA TATGGTGATG GCTTCTGGGT AGGGAAAGAA TTAGGGTTGG GAATGGGATG
GGTTTGGAAT TGTGACTGGG ATGGGGACAG GCATGGGATT GGAGACCAAG AGGGAGTTGA
GGATGGTTTG GGGACCGGGG GTGGGGATGG GGGTGGGGCT GGGGCTGGGT GTGGGGTTGG
GATTGGCGTT GGACGTGGAG ATAGAGATCA GGGTTGGTGG TGACCTGCCC CATCTTCCTC
AGAGTTTTCA GAGTCCGTCT CGGCCACTAC TCCCTGTCAC CAGTTTATGA ATCTGGGCAG
CAGATGTTCC AGGGGGTCAA ATCCATCCCC CACCCTGGCT ACTCCCACCC TGGCCACTCT
AACGACCTCA TGCTCATCAA ACTGAACAGA AGAATTCGTC CCACTAAAGA TGTCAGACCC
ATCAACGTCT CCTCTCATTG TCCCTCTGCT GGGACAAAGT GCTTGGTGTC TGGCTGGGGG
ACAACCAAGA GCCCCCAAGG TGAGTGTCCA GGTTCTTCTT GATACCGACC CATCTCTGCC
GCCTTCCATC TTTCTCCACT TCTCATTGTG TTCCTGTTTG ACAGTGCACT TCCCTAAGGT
CCTCCAGTGC TTGAATATCA GCGTGCTAAG TCAGAAAAGG TGCGAGGATG CTTACCCGAG
ACAGATAGAT GACACCATGT TCTGCGCCGG TGACAAAGCA GGTAGAGACT CCTGCCAGGT
GAGGACACCT CTCTTTATTC AGCAGATACA CACTGAGTGC CAACTCGGTA ACATGGAGCG
TTGCCAAATT CTGAGAATCC AGCAATTGCC AAGACAGTCA GGACCCCTGT TCTCACAGAG
CTCATACCCT AGAGTAGTGG TGTTTAGTAG AAATAATGCT GAGCTGCTTA TGTCATTTGC
AGTTTTTTAG TAGCCACATT AAAACAGGTA AAAAAGGCTG GGCGCAGTGG CTCACACCTG
TAATCCCAGC ACTTTGGGAG GCTGAGGCAG GCAGATCACC TTTGGTCAGG AGTTTGAGAC
TAGCCTGGCC AACATGGCGA AACTCTGTCT CTAAAAAAAA ATACAAAAAT TAGCCTGGCA
TGGTGGCGGG CGCCTGTAAT CTCAGCTGCT CAGGAGGCCG AGACACAAGA ATCACTTAAA

FIGURE 3 (cont'd)

CCCAGGAGGT GGAGGTTGCA GTGAGCTGAG ATCGTGCCAC TCACTCCAAC CTGGGAGACA
GAGTGACACT TTTGTCTCAA AAAGAAAAAA AAAAACAAGT AAAAAAGAAA CAGGTGAAGT
TAACTTTAAT AACCCAATGT ATCCCAAATA CAATCATTTC AAAGTGTAAT TAATATAAAA
CAATTATGAA TGAGATACTT TACATTCTTT TCTTGTTTTC ATATTAAGTC TTTGAAAGTG
AGTATATATG TTATGCTGAC AGCACATCTC AATTTGGACT AGCTACATTT CAGGTGCTCA
GTAGCCACAT GTGGCTAGCA GTTACTGTAT TGGATGGCAC GGATCTAGAG GGAAAGATCA
GGGCTGTTTT GTATGGTTGG GCAGGTTGTG CACTGCATAA AGATACCATA TCTAATAGGG
GCACTCCGTG TTACAGATGT CAGTTTTGGC AGTTTTCAGG CGTGTGGTAG TTAAGTGTCT
TGTTTCAACA AAATCTGTAA TATGACAGTT TTCTAGCAAG TGCTGGTAAA ATATCTTGAG
GAAGGAAAAG AGAAATCTGG TAGGTATTTT TACAAGAGAA TATTTAATAC AGGGGATTAA
TTGCAAAGCT GCTGGAAGGG CTGGAGGAAC AAAGTTAAAA AATAAAAAAC TCTGTGGTCA
AGAATCTGCA TAAATAGGGC AATTTCAGAG AGTGGTAAAG GTTAACCCCA AAATAAAACA
TGGTTTTAGG ATAGTAAACA ATAAGGGCCA ATATTCAAAA AGGTGGTCAG GGGAGCCTCC
TTGGAGAGGT GGCATTTGAG CAGAGAATGG ATGACACAAA GAAGCTAAAC TCGTGAAGTT
TAAGGGGAAA GAAAAGGCAC GTGCAAAGGC CCTGAGGCAG TAAGGAATTT GGCTGATTCA
AAGAAGAAGA GGAAACCAAT GCAACTGGAG AACAAAAGTG GGGGCAACAG TAGAAAGTGA
CGCTGGAGGT GTAGGCAGGG GCGAATGCTC TGCAAGTATT TCTTGGTCAC CAACACAGAG
CTTCCCTATG TTCTAATGGA AGCTGTATCT GTTGAGGAAG ACAGAATTTA AAATCAAACT
GTTACATCAA CCAGCACCCT TCTCTGTATT CAGGCTCCCA AGGGATCTAG AAGGACGTAA
GTTAACAAGC TCTCATTAGC AGGGTGTGTG TTTCAACAGT AGTTAGGAAG CTGGGGATTC
AGGAGTACTC CAGTCCCATG GCTATGAAAA GCTCCCCCCA AATTGTACAA ACCTGACAAA
TGCAACACCT CCCCAGCTCT CCCCATTTCT TCTCTGTGCC CTGGGTGTGG GGGGGTGGGT
TGCGAGGGGG AAAACTTTTA ACAGAAGAAA GCACATCTCG GCCGGGCGTG GTGGCTCACA
CCTGTAATCC CAACACTTTG GGAGGCCGAG GCGGGTGGAT CACTAGGTCA GGAGATGGAG
ACCATCCTGG CTGACACGGT GAAACCCTGT CTCTACTAAA AACACAAAAA ATTAGCCGGG
CGTGGTGGCA GGCGCCTGTA GTCCCAGCTA CTCGGGAGGC TGAGGCAGGA GAATGGCCTG
AACCCGGGAG GCGGAACTTG CAGTGAGCCG AGGTTGCACC ACTGCACTCC AGCCTGGGCA
ACACAGTGAG ACTCCGTCTC AAAAAAAAAA AAAGAAAAGA AAAGAAATCA CATCTCATTC
AAGTGGTGGC ATTTAAAACT ATTTAGCCTT TCTGTAGGCA AGGTTAGTAT CTTGTTTTTC
CAGACCTCAA GGTGTTTTTT TGTTTGTTTT TTCATACCGG TGTGTGGTCT GGGTGTGGCC
ACTAAAAGCT ACAAGCAAGA AATAATAACA ACTACAACAA TACTAATACC AATAGTATAA
AAATAATAGC ATCTGGCTAA TTGCTGGACA CTGTTTTAAG TGGTTTGCAT GCCTCAGCTC
ATTAACTCAT TTACCTGTTA TTATTGGCCC TATTTTACAA ACAAGGAGCC AAGGCTCAGA
GCAGTTAACT AACAGCCTCT CAAAAGAAAC TCTGCAGAGA TATTAAATTT AAAAAATAAT
GAGAGAAATT AAACCACAAG AAAGTTGAAA TTTAGAGGTA CAGGCAGCTA AGCTTGTTTG
CTTTGAAACA GTGTCTGCTA CTGGGAAAAA GGCAAGTCTT GGCTTTCCTA ATAATTGATA
CCAGGACTCT GTAATTCATA TTTTGCATGC ATGTAAGTAA GAAATGAAGC CGGGTGCAAT
GGCACATGCC AGTAATCCCA GCACTCTGGG AGACTGAAGT GGGAAGATCA CTTGAGCTCA
GGAGTTCAAG ACCAGCCTGG GCAACTAAAA ATTAAAAAAA TAAAAATACT AATTGTTTTT
ATTTTAGTAG ATTTTATTCA TACCACTTAC ATCATTATTG TAGTATGTAC ATATTTATTT
CTTTTCTTTT CTTTTCTTTT CTTTTTTGAG ACGGAGTCTC GCTCTGTCAC CCAGGCTGGA
GTGCAATGGC ACCATATCAG CTCACTGCAG CATGCGCCTC CTGGGTTCAA GCATTTCTTC
CACCTCAGCC TCCCAAGTAG CTGGGATAAC AGGCACCCAC CACCATGCCT GGCTATTTTT
TTTTTTCCGT AGAGATGGGG TTCCACCATG TTGGCCAGGC TGGTCTTGAA CTCCTGACCT
CCAGTGATCT GCCTGCCTCG GCCTCCCAAA TTGCTGGTAT TACAGGTGTG AGCCACCGTG
CCCAGGTGGG AGATAGACAT TTCTCTCTAC CTCAAACAGA GGTCCACTCA AGCTACTTTT
CATTTTCTTC ATAAATATTA GCCGAGTGGC TATTTTGCAC CAGGAATGGT TCCAGGTGCT
GTGGATATGG CATCAGGCAA AACAGACCAA AAACTTCCTG CCGCGTGGAC CTCATGTTCC
CCAAGTGGAA GACAGGCAAT AAAGAGATAG ATAAATATGT AGTAAATTAA AAAAAAAAAA
AATTAGCCGG GTGTGGTGGC TTGCACCTGT AGTTCCAGCT ACTTGGGAGG CTGAGGTGGG
AGAATTGCTT GAGCCCAAAC GTTTGAGGCT GCGGTAAGCC ATGACTGCAC TGCTGCACTC
CAGACAGCAG CCTGGGTGAC AAAGCAAGAC GTTTTTGTCA GAAAGAAAAA AAAAAGAGAC
GAAGGGAGGA AGGAGAGAGA AAGGAAGGAA GGAAGGAGAA AGAAAGGAAG GAAGGAGAAA
GAAAGGAAGG AAGGAAGGAG AAAGAAAGGA AGAAAGAGAA AGAAAGAAAA AGAAAGAAAG
AAAGAAGAAA GAAAAGAGAG AGGAAGGAAG GAAAGAAGGA AAAGAGGGAA AAAAATGACT
GTTGAAGAGC AGTGAGTATT ATTATAGGAG GGTAATTATA GGGAGGTATG GGGAATTGAA
GACAGGAAAC ACAAATTAGT CCAAGCGAAT GGATTTCTAT TGGGAGTGAT TCTGCCCCTA

FIGURE 3 (cont'd)

GAAGACACTG GCAATACCAG GAGACATTTT TGGTTGTCAC AACTATATGG AGGGGCATTA
CTGGCAACTA ATGGATAGAT GCGAAGTGTG CTGTTCAACA TGGTATGATG CACACGGCAG
GCCTCCACAA CAAACCATTA TCCAGCTTCA GATGCCCACA GTGCCCAGAT CGAGGAACCC
TCATCCAGGG GCTGAGAACC GTATTTTTGC AGAAGGGAGG TATAAGGATG GGTTGGTGGA
GAATGGGGAA GGAAGGTGTG TGTCCAGTAA GAGAAATAAG GCCTGCACAG GCTGGAGGGG
AGAGTGAGAG AGAAAGGGAG GCGGAGAGAT ACACGATGAG GGAGACAGGC TGGAAGAGAA
AGTAGAGACG AAGATTCGAG ATGTGGAGAG GAAGGGTCAC AGACCCCCCC GAAATGATGT
GTGGACAACA GGAATCTGGA AGAGGAAGAT GGAGTGGAGA GTGACAAATG GGGTCTAAAG
GTTGAACTTG GAGGCCAGGC ATGGTGGCTC ACGCCTGTAA TCCCAACACT TTGGAGGCTG
AGGTGGGCGA ATCACTTGAG GCCAGGAGTT CGAGACCAGC CTGGCCAACA TGGTGAAACC
CCGTCTCTAC AAAAAAAATA CAAAAAATTA GCCGGGTGTG GTGATGGACA CCTGTAGTCA
CAGCTACTTG GGAGGCTGAG GCAGGAGAAT TGCTTGAACC CGGGAGATGG AGGCTGCAGT
GAGCTGAGGT CAGGCCACTG CGCTCCAACC TGGGCAACAG AGTAAGACTC CATCTCAAAA
AAAAAAAAGC TGGATTTGGA GTGAAATATT AATAACATTC TCCCTCTCTC TCCTTTTGCC
TGTGTCTCCA TCTCTGTCTT TTTCTGCATT TCTTCATCTC TGTACTTTCC ATCTCTGTGT
GTCTGTTCCC ATCTGCTTCT CCATCTATGG GCATCTCTGG GTCTCTCATG TCTCCTTCTG
CCCACTTTGC CACATCTCTG CCTCTCTCAT GCCCCCCTTT CTCTCCTGCA GGGTGATTCT
GGGGGGCCTG TGGTCTGCAA TGGCTCCCTG CAGGGACTCG TGTCCTGGGG AGATTACCCT
TGTGCCCGGC CCAACAGACC GGGTGTCTAC ACGAACCTCT GCAAGTTCAC CAAGTGGATC
CAGGAAACCA TCCAGGCCAA CTCC*TGA*GTC ATCCCAGGAC TCAGCACACC GGCATCCCCA
CCTGCTGCAG GGACAGCCCT GACACTCCTT TCAGACCCTC ATTCCTTCCC AGAGATGTTG
AGAATGTTCA TCTCTCCAGC CCCTGACCCC ATGTCTCCTG GACTCAGGGT CTGCTTCCCC
CACATTGGGC TGACCGTGTC TCTCTAGTTG AACCCTGGGA ACAATTTCCA AAACTGTCCA
GGGCGGGGGT TGCGTCTCAA TCTCCCTGGG GCACTTTCAT CCTCAAGCTC AGGGCCCATC
CCTTCTCTGC AGCTCTGACC CAAATTTAGT CCCAGAAATA AACTGAGAAG

FIGURE 4

KLK-L 3

CTTGAACCCA GGAGGCAGAG GTTGCAGTGA GCTGAGATCG CGCCACTGTA CTTCAGCCTG
GGTGTCAGAG CAATACTCCG TTTTGGAAAA CAAACAAACA AACAAACAAA CAAAAAACAG
ATGGAGCAAC TGAGAGAGGT CTTGTGACTT GCCCAAAGTC ACACACCTCA TCACTAATCA
CACCTAATCA TTGAGATTTG GACACACATG GTTCAGTTCC AGAGTCCATG CTCCAAACCA
TGACGACACA GTGAGAGAAC ATTCAAGGGG AGCCCAGACC CAGCTTCATA ACCAGGCCTG
TGAGCAGGAG AAAGTGGAAG GGATCGTAAG TGCCCAGGGG AGGCAAAGAT GGACTCTGCC
TGAGGATCTC AGAGATTTCC TGGAGGAGGG AGAATTGAGG TTGGGTGTTG AAGGATGAGT
GGGAGTTCAC CAGGAAAAGA AGGATATGGA GAAAGACATT CACTCATTCA ATGAACATCT
CCTGAGGACT TCTGCAAGCC CTGTTCCGCC TGGAACGGGG TGATGCTGGG ACACAGAGAT
GAGTCAGACC TGGGCCCAGC CCTCCAGAAG CTGTCCACCT GGTGAGAAGG AATGATGAGG
AGAGAGGCAG GGAGGATGGG GTGATGGAAG GGACAATGGG GTGGGGGGCA GGGAGATGGA
TGAAAAAAAT ATATAGCAAA TGTTCTCAGG ATTTGGCAAA GATCAGGATG TATTAAGAGA
GAGCACAGGG CACTTGCTAC CTGGAAGGTT GGGCACCTGG GTCCTTGGGT GGTGGAGCCG
TGGGGAAGGG GGCAGGTTAT GACAAGAGTG GGTTAATCCA GATGGAACCA GATTTCTCAA
CATTCTAGGA GAGGGCCTTG TCCTTGTGGG AAGAGGCCCA AATCCCCAGG GCAGGGAAGG
TTCTGCAAGG TGTGTAAACC TGTGCAGCTG CCTGTGGTCT CTGCCTCACT CCACCTGGAT
TTCCCTCAAT CTTTCCCGTG TTCTGTCTCC TCCTCCCACT CCTCCTCTCA TCTTGGGTCC
TTCTGTGCCT GTACCTCCCT CTCTTTGTAT CTTTTGCTCT TGTGTCTGAG TCCTGACTCT
GTCTTCCACC CCTCGCCTCC TTTCTGGGTG GTCCCCCTGC ACATCCCTCC AGCCTGCCGT
GGGAGGTTGG TCTCTGCACA CCACTGCTTT ATCCAAAATA AACCTGCTGC ACCCCAGGAC
CTTAGGCTTC AAGGATCTCC CTCCTTTTCC AGGACACAAA AGATTCTGTA TCTTGTAGCC
TAAGGTGATG AGGAATGAGG TCTCCCACTC TGAAGACCCC AGAGGAGGTG CCCACAACCT
CTCCACACCC CCAGCACTCC TCCTCCATTC AGTCAAGCTC TGGCCCAGCA AGCCGCCAGT
TCATCCCAAA AGGGGGGTCC CCCTGCACTT ACCTCCTCTC CCAAGGCCCC TGTCACAGCC
CCAGGGCTTC CCCCTCCCCC AGGTACATTT CCCAACCCCG ATTAATCACA GGGGCGGCCC
CATGGAGGAG GAAGGAGATG GCATGGCTTA CCATAAAGAA GCACTGGACG CCGGGTGCAC
GTTCCAGGAT CCAGGTGCCC AGGGGTCATG AAGCTGGGAC TCCTCTGTGC TCTGCTCTCT
CTGCTGGCAG GTGAGGCTCC CAGGCTGGCT GCCCCTTCAC GGCTGTACTA AGGTCACCTT
GCTCTTCCCT CCCATCCCAG GCTTCTGCCT CCTGCCCTCT AGGCTTCTCA GCATCCTCTC
CCTGCCCTCC CAGCCTGCTC TTCGCTGACC CCTTTGTCCC TCATCCCCAC CCCAGGGCAT
GGCTGGGCAG ACACCCGTGC CATCGGGGCC GAGGAATGTC GCCCCAACTC CCAGCCTTGG
CAGGCCGGCC TCTTCCACCT TACTCGGCTC TTCTGTGGGG CGACCCTCAT CAGTGACCGC
TGGCTGCTCA CAGCTGCCCA CTGCCGCAAG CCGTGAGTGA CCCAGGCTGG CCATGCTGGG
GAGGGACAGA GGCTGGGGGT CAGGAGAGGG TGAGGGGTGC TTTAGGCCAG AAGTGCGGAG
CCTCCACTTC TGATACCACA AGTTCAACTC TTAGAAGTAG GAAGGGTAGC CTCCCAAATC
CTAAAATTCT AGAGACCAGC AATATCTCAT TTGAGAAGTC TAAGATTCGA AACTTAGGCT
CTTCGAATCC GAGACTGACC CAGAGAAATC CAGAATCGTA GAATCCTAAA ATCTTGAATT
TATGAAATTC TGCAATAGCC TCAGCAAATT TTAGAATCAT AGATTCGCAG ACTATTAGAA
TCTTAGCAGT CTGGGTCAGC ACTGCCCAGA GGAATTATGA TGCCAGCCAC ATGTGTAAGT
TTAAATTTCT GGTGGACACA TTTAAAAAAT AAGGAATGAG TAAAATTAAT TCTAATAGAT
TTAACTTGAC ATACCCAAAA ACTTATTTTG ACATGTAATC AATTTTTAAA TACGTATGAA
CGATACAGTT TACTTTTGTT TTGGTACTAA GCCTTTGAAA TCTGTTCTGT ATTTTACACA
CATAGCCTGT TACAAAATGG ACTAGCCACA TTTCAAGTGT TCAATAGCCA TAATGGCTAG
TGTGATCCTA GAATCTTAAA TTCAGAGCTT TCTAGATTCA TTGAATATTG AAACTCACAG
TACTAGAATC TTTGATTCAC AGTATCCTAG AATATTGAGA TTCAGATAAT TCTGTAGTCT
TAAACTATTT GAATCCCAGA CTCTTAAATT TCTAAGGTTA TAGATTTATA GAATGATGAC
ATTCTAGTCT TTCTTTTTTT TTTTTTTTTT TTTTTTGAG ACAGAGTCTC CCTCTATCTC
CCAGGCTGGA GTGCAGTGGC ACAATCTCAG CTCACTGCAA CCTCTGCCTC TCGGGTTCAA
GCAATTCTCC TGCCTCAGCC TCCTGAGTAG CTGGGATTAC AGGTATGCAC CACCATGCCA
GGCTATTTTT TTTTTTTTTT TTTTTTTAGT AGAGACGGGG GTTTCACCAT ATTGGCCAGG
CTGGTCTTGA ACTCCTGACC TTGTGATCTG CCCGCCTCGG CCTCCCAAAG TGCTGGGATT

FIGURE 4 (cont'd)

ACAGGCGTGA GCCACCGCGC CCAGCCAAAA TTCTAGTCTT TTTGTCCTAG AACATTAAAA
TTCTATGTTC AAATCTTAGA TTTAATTCAG ATAATGTTAG AATCCTGGAG TTTTTTTGAT
CCAGGGGAAT CTGGAATGTT AGAATCTTGG ATTCATAAAA CTCTAAACCT TGAGCCTCTA
GATTCTAGAA TCATGGATAA TAGTGTGTCG GAATCTGAGA ATTCTAGAAT CTTAGGTTCT
GGGCATTCTA ATAGTATCCT GGAATCCACC TGATGCAGGA ATCCTCTCTC CATTGCCTCT
GAAAAGTGAC CATCCATACT GTTCCAATTT TCTTCCCTCC ATGAGTAAAG CACTGATTGT
GGTAAGAGAT GCTGTGTGGG AATTTCCCAT CATGCATTGC TCCATGATGG AACCTCCTTT
AACTTAAGCC TATACATCAG ACTGGGAGAA CGATGTTCAG ATTTCAGCCG AAAGTGAAGC
AGGAGAAATG CAGAGATATG AAGGTGGAAG AGAGTGAGAG GCAGGGGAAG GGTAGGGGGA
TGAAGGGATG TAGGGGTGAG GACTACTTTT CCAGATCCAG AGCCAAGACA GCAAGAATGA
CAGAGAGAGA CAGACACAGA TGTTTCTGGT TCCCCAACCC TGAATTCGCA GTCATTAGCC
TGCTGCCTAA TGTCAGAGGT CAGAGGCTGG GGAATGGACT TGTCATCCCC GAAAGGATCC
CAGCTGTCTA GGGCATGGAC CAGAAATGAA ACAAGTGCGC TGAGACTGTG GTGAGGGCTT
AAGGTTAGAC ACCAGGAAGA CATGCATTGA AGGGTGAAGG ATATGATAGA CAGGAAAAGC
TGAGGCCAGA GATGACCCCC AATTTGGGGA TTTTCCATAT CCCATCCCCT TTCATACACA
CGCACACGTA TACACACACA CCACTTAGAC ATACAGAGCC GCTCCCACAG AAGCCACCAG
ACCTGTGGGG GCAGGGGTGG GGCGGTTGTT ATGTGGTAGG TGGGGTCCCC CGTGCCCACA
CCGTTCCTAG GGACCCAAGT CACCACCAAG GCTCCAGGTG AGTAGGGAGG AAGGTGGCTC
ACTCAGCCTG GGACTAGGAG CGGGGGCTTT GTGGGGAGAG CTACAAAGAT GGAGACACAC
AAAACATCAG AGTGGGGACC AGGGACCCAG AGGAGGTGTG TGCCTCGCTT AAAATCACAG
TACCCTGGGC CAGACATAGA TGATGAGGGT GCAGAGAGGG TGTGTGGCTT GCAGAGGGTC
ACACAGCACC CTGATGGACA GGAAAAGAGG GCTGGGGCTG AAAGGACTTT TACCTTTCCC
CCAGCTTGAC CTCTGAGGCC TGTCCCAGCA GGTATCTGTG GGTCCGCCTT GGAGAGCACC
ACCTCTGGAA ATGGGAGGGT CCGGAGCAGC TGTTCCGGGT TACGGACTTC TTCCCCCACC
CTGGCTTCAA CAAGGACCTC AGCGCCAATG ACCACAATGA TGACATCATG CTGATCCGCC
TGCCCAGGCA GGCACGTCTG AGTCCTGCTG TGCAGCCCCT CAACCTCAGC CAGACCTGTG
TCTCCCCAGG CATGCAGTGT CTCATCTCAG GCTGGGGGGC CGTGTCCAGC CCCAAGGGTA
TGACCTGGCC CAGAACTCTC TCTGAAACTT GCTCCCTCAC CCCTCTGTCT CTGCCTTTTC
ATCTCTGTCT TCTCCTTTTC TCTCTCCTCT CTCTCTCTGT CAGTCTATGT ATCTGCCAAT
CGATATATTT AACCAAATAT AAGATGCTAG CATTTTTAAG ATGTGCCATT ATTTCATGAA
CTGCGAAGAA GTGGAAGAAG GAGGAGGAGG AGAAGAAAAA AAGGAGGAGG AGGAAAGATC
CCATTAGATC CCATTGATTA TATAACACCA TTTTCTGGAA GACACATTCT AATTTCAGAG
TGTTTGTTTG TTTGTTTGTT TGTTTGTTTT TGAGACAGGG TCTCGCTTTG TTGCTCAGGC
TGGAGTGCAG CGGTGTGATC ACGGCTCATT GCAGCTTTGA ACTCCTGGGC TCAAGTGATC
CTCTCGCCTC AACCTCCCAA GTAGCTGGGA TTACAGATAT GCACCACCAC ATCCCACACC
GGGGTCATTT TTTTATTATT TATTATTATT ATTATTATTA TCTTTTTTTT TGTATTTTTA
GTAGAGACAG AGGTTTCACC ATATTGGCCA GGCTGGTCTC AAATTCCTGA CCTGGTGATC
TGCCCGCCTT GGACTCCCAA AGTGCTGGGA AAACAGGCAT GAGCCACTGC ACCCAGCCAA
AATTCTAGTC TTTTTTAAAT CTAGTCATAT CTTAGATTTA ATTCAGATAA TGTTAGAATC
CTGGAGTTTT TTGATCCAGG GGAATCTGGA ATGTTAGAAT CTTGGATTCA TAAAACTCTA
AACGTTGAGC CTCTAGATTC TAGAATCATG GATACTAGTG TGTCAGAATC TGAGAATTCT
AGAATCTTAG ATTCTGGGCA TTCTAATAGT ATCCTGGAAT CCACCTGATG CAGGAATCCT
CTCTCCATTG CCTCTGAAAA GTGACCATCC ATACTGTTCC AATTTTCTTC CCTCCATGAA
TAAAGCACTG ATTCTGGTAA AAGATGCTGG GTGGGAATTT CCCATCATGC ATTGCTCCAT
GATGGGACCT CCTTTAACTT AAGCCTTATG CTAAAAATTT TTATTATTTT TAGCAAAGAT
GAGGTCTTGC TATGTTGTCC AGGCTAGTCT CAAACTCCTG GCCTCCCAAA GTGCTGAGAT
TACAAGTGTG AGCCACTGTA CCTGGCCCAG AGATGTTTAA ATGTGAAATG CGTTCATCTT
AGAATGGGAA TAAGACCATG TCTCTCAGAG TCACGGATCA CTGACCCATT AGCCAAATTG
GGTCAGTGGA TTGGAAAAAC AGTCTGAATT TGTTGCTGCC AATATCTAAA ACTTGGAAAG
TTTTATACAA AAGCCAGGTT TCTGGATTCA CCTGAAAAAG TTTGAAGAAC TCACATTCCC
AAAATAGCAA GCATTGGGCT GAGTCAATGG AGGCTGCCCC CTTCAGCCAA GATAAGTTCT
CTGATTCACT CCAATGGACC CAAATGGCTC CTGTCTCCCT GCACAGCCCC CGTCCCGGAG
TTCTGTTTAC CAATTCTGTT TATCATATCC CTTGATGCAT CGGAGCCTGC AGCCATGTCT
TATATAGATG CACATGTGTA TTATATATCC ATATCCACAT CTATACTGAC TACACTGTAT
CTGGTATCTC TGTCTATGTC TCTGTCTCCA TCAGTGACCA TCTTCCTGCA AATCTCTTTC
CTTTTATCTC ACTGCCTTCA TTCCACCCCT TGAGGTCTGG GTCTTTTTCT ATTTCTTTTT
TTTTTTTTTT TAAGAGACTG AGTCTTGCTC TTGTTGCCCA GGCTGGAGTG CAGTGGTGTG

FIGURE 4 (cont'd)

ATCTCGGCTC ACTGCAACCT CCACCTCCTG GGTTTTAAGT GATCCTCCTG CCTCAGCCTC
CCGAGTAGCT GGGACTACAG GTGTGCAACA GCATGCCCAG CTGATTTTTT GTATTTTCAG
TAGAGACGGA GTTTCACCAT GTTGGCCAGG ATGGTCTCAA TCTCTTGACC TTGTGATCCG
CCCGCCTCAG CCTCCCAAAG TGCTAGGGAG TTATATATGC ATCTCCTCTT ATCTCTTGGC
TCTCTGCATG CATCTTTCTG TTTCTCTTCC TTCCTTTCTT TTTTTTTTTT TTTTTTTTTT
TTTTTTTTTT TTTTTTGAGA CGGAGTCTTG CTCTGTCTCC CAGGCTGGAG TGCAGTGACC
AGTCTCGGCT CACTGCAACC TCCACCTCCC AGGTTCAAGT GATTCTCGTG CCTCAGCCTC
CCGAGTAGCT GGGATTACAG GCGCCTGCCA CCATGCCTGG CTAATTTTTG TATATTTAGC
AGAGATGGGG TTTCACCATG TTGGCTGGGC TGGTCTCAAA CTCCTGACCT CAAGCGATCC
GCCGGCCTCG GCCTCCAAAA CACTGGGATT ACAGGCATGA GCCACGGTGC CCGGCCAGCC
TCTCTCTCTA CTTGGCCCTC TTCCTCCTTG TCTCCATTTG TTTCTCTTGT GTGCTATGAC
TGTCTGTCTG TCACTGTCTC TTGTCTCTAT CTTTGAGAGT CCTAAATGTG GCTCCATTGG
TCCTTTGGAA AAGCTGCAGG GAGGACTCAG GGCAGTGGGG TGCTGAGTGT GTTGGAGACA
GTTGCAGATC CTTGACAGTT CTCTTCCCTG ACAGCGCTGT TTCCAGTCAC ACTGCAGTGT
GCCAACATCA GCATCCTGGA GAACAAACTC TGTCACTGGG CATACCCTGG ACACATCTCG
GACAGCATGC TCTGTGCGGG CCTGTGGGAG GGGGGCCGAG GTTCCTGCCA GGTGAGACCT
TACTCTGGGG AAAATGAGGC TGTCCTGCCA AGTTTTCTAG GATTTAGGGG AGCAGAGGGG
TCGGCCCCCA GCCTTCCTGG GTCAAAATGA GAAGGAGACT GGGATACCTG GTTCCTGGGA
GAGGACGGGA CCAGGGCCTG GACTCCTTAG TGTAAAAGAG AAAAGGTCTG GAGGTCCAGA
CTTCTGGATC TACAGGAGGA GTGGGCTGGG CGTCCAGAGT CTGAGTCCTC GGGGAGGAGG
AGGTTAGGTC CTGCGGGGAG GTGGGCCCTC TGAGCTTTTA CTCCTGGGTC TGAGGAAGAA
GAGGCTGGAG ATGGAGGACT CTCGGATGTT GGAGGAGGAA GGGGCTGGGG CCTTTCTGGG
AGGGAGGAAG TGGCCCGTGT AATTGTCATG AACAGAGTGG CCTAACAGTT CCTCTGCCCT
TCTCTCGCGT ACAGGGTGAC TCTGGGGGCC CCCTGGTTTG CAATGGAACC TTGGCAGGCG
TGGTGTCTGG GGGTGCTGAG CCCTGCTCCA GACCCCGGCG CCCCGCAGTC TACACCAGCG
TATGCCACTA CCTTGACTGG ATCCAAGAAA TCATGGAGAA CTGAGCCCGC GCGCCACGGG
GGCACCTTGG AAGACCAAGA GAGGCCGAAG GGCACGGGGT AGGGGGTTCT CGTAGGGTCC
CAGCCTCAAT GGTTCCCGCC CTGGACCTCC AGCTGCCCTG ACTCCCCTCT GGACACTAAG
ACTCCGCCCC TGAGGCTCCG CCCCCTCACG AGGTCAAGCA AGACACAGTC GCGCCCCCTC
GGAACGGAGC AGGGACACGC CCTTCAGAGC CCGTCTCTAT GACGTCACCG ACAGCCATCA
CCTCCTTCTT GGAACAGCAC AGCCTGTGGC TCCGCCCCAA GGAACCACTT ACACAAAATA
GCTCCGCCCC TCGGAACTTT GCCCAGTGGG ACTTCCCCTC GGGACTCCAC CCCTTGTGGC
CCCGCCTCCT TCACCAGAGA TCTCGCCCCT CGTGATGTCA GGGGCGCAGT AGCTCCGCCC
ACGTGGAGCT CGGGCGGTGT AGAGCTCAGC CCCTTGTGGC CCCGTCCTGG GCGTGTGCTG
GGTTTGAATC CTGGCGGAGA CCTGGGGGGA AATTGAGGGA GGGTCTGGAT ACCTTTAGAG
CCAATGCAAC GGATGATTTT TCAGTAAACG CGGGAAACCT CA

FIGURE 5

KLK-L 4

ATTAAGAAGG ACCCAGACAT ACAACCTCTA AATTCTGAGG GTCATCCAGT AGAATATTCC
ATATATGTAT ATATGAAATA TCCTATATCT GTGCTGTCCA ATTATCCAGT AGCCCCTTCA
GGCTATTGAA CATTTGAAAT ATGGCTGGTG TGACTTAAGA ACTGAATTTT TAATTTAGTT
TTACTTCATT TTAATTAGTT TAAATTTAAA TAGCCACATG TAGCTAGTGG CTACCATATT
AAACAACATA GGTCTGGAGA AAGGACTGTG CAGAGAGAGG AAATAGCAAG TATAAAATGT
CTAGTATGGG GGCATCCAAG ATGATTTAAA TTCTTCTTTT CTTTAAATGC CTGGTGTGTT
TGAAGAACAG GCCCATGAGG CTGGACTAGA GGAAGTCAGA AGAAAGAGGT TGGAGATGGG
GTCAAAGAGG CTGGCAAGGG CCAGACAGCA CAGAGTCCTG CACACCTTGG GAAGGCTTTT
TGGATTTTAT TTTAAAGAAA GTTGAGCCTG GGAACAACAT CTGACTTTCT TTGTTTGAAG
AGTCCTCAGC CTACTTTGAG AAGACTGGAT CGGAGGGATG TAAAAGTGGA AGGATTTAGG
TTAATGTTGT AGTCATTTGG GCTACAGAAG ATGGGGCATG GACCAAGATG GTGGCAGAAG
TGTGGAGATA ACTGGATATT TGGGAGATAA AACCAATAGG AACTGGTTGT GAGTGATGAA
GGAAAGAAGA GAAGCAAAGA TGACTCCCAG GTTTGGGGCT GAGCACTGAG GTGGGAAATA
CTGGAGCGAA CAGTTTTGAT TGAGAAGAAT CAAGTTGGGA ATACAAAGCT TAAGATGCCT
GTAAGGCATC CAAATCAACA GTGTTTGAGT TTTGAGCTTA AAGAAGAGTT CAGGGCTGGA
GATGATTAGC CTATAGCTGG TATTTAAAGC CATGGAGGCA ACCAGTATAT ATGCAGTGAA
AGGATAGAGA GATGGGTGGA AAGATGATTG GATGGATGCA TGGATGGATA TATGGATAGA
TGGATGGATG GATGGTTGGA TTGGATGGAT GGATGGATGG ATGGATGGAT GGATGGATGG
ATGGATGGAT GAATAAATGG ACCAGTGGAT GGAGGGACAG ATGAGTGGAT GGATGGTTGG
ATGGATGGAT GGATGGATGG ATGGATAGAT GGTTAGATGA CTACCTAAAT GGATGAATGG
ATAGATGGAT GAGTAGACGG ATGGACAAAT CAATAGGATG AATGGGGGAT GGATGATTGG
ATAGATTGAT GGATAGATAT TGCCTAGGTG GATGTGTAGG TCAGTCTCAC TTCTACCTCC
TGAAATCCAT CTTCTGGTAG AATGATATAA AAAATGCATG TGGAGAGAAA GTCAGGCTCC
TGCTTACCTA TCAGCAACAT CCTCATTTTG TGAACTCTTC TGTTAACCCC CAGTGGAGGA
TTTGGTACTT CCTGAGAAAA TAATGTCACC CCTTTGCCCT AATTCATCTC CACTTGGTCA
AGAATAGCAA CTGCCATAGG TCGGCAAATT CATCTTCAGT TCCTGGTCAC CCAGGGCAAT
AATCCGACCC TTACCCCAAA CCCAGAAACC ACAAGCCCAG GGCTCCTCTG GCCCCCTGGAT
CCCAGTTTTC TAACAATCTC TCTTCTTTAC CAGGTGTCTC CCAGGAGTCT TCCAAGGTTC
TCAACACCAA TGGGACCAGT GGGTTTCTCC CAGGTGGCTA CACCTGCTTC CCCCACTCTC
AGCCCTGGCA GGCTGCCCTA CTAGTGCAAG GGCGGCTACT CTGTGGGGGA GTCCTGGTCC
ACCCCAAATG GGTCCTCACT GCCGCACACT GTCTAAAGGA GTATGTGGGG GCCGGGGGAG
CATGGGGTAG GGATGAGAAT GGGACTGGGA TTGTGGATGG GGTAGAGTTG GATTTGAGGA
TGGAGTTGGA GTTAGGGTTG GGGATGGACA TGGGAGTGAG AATGAGGTTT GGGGTTGAGA
TATGGGGATT GGGTATGGGA ATAGAATCAA AGTAGGGGAT TTGGATGGGA TTGAAGTTGA
GGATGGGGGA GATGTATTTG GAGATGAGGA AGGTAGGATG GAGAAGAAGT TAGGTTGGGG
ATGGGAAGAG GTTGGGGCTG GGATGGGGAT GGAAATGGGC TCATCTTCTT TCCTAACCAC
CTTCTTTCTG CACCCACAGG GGGCTCAAAG TTTACCTAGG CAAGCACGCC CTAGGGCGTG
TGGAAGCTGG TGAGCAGGTG AGGGAAGTTG TCCACTCTAT CCCCCACCCT GAATACCGGA
GAAGCCCCA CCCACCTGAA CCACGACCAT GACATCATGC TTCTGGAGCT GCAGTCCCCG
GTCCAGCTCA CAGGCTACAT CCAAACCCTG CCCCTTTCCC ACAACAACCG CCTAACCCCT
GGCACCACCT GTCGGGTGTC TGGCTGGGGC ACCACCACCA GCCCCCAGGG TATGCACCCA
CACAGGTGGC CTGAGGCCCC ATAGGAGTGG CTGGGGAAAC AGGGGCAGAG ATGGGAGGGA
AGGTCTGAGG
TAGGTTCCTT TATATATAAA AATATAAATA AGTAAATAAA TATATATATT TAAAGTTAGC
TGTATCCTTT ATATAAATAT AAATTCATGA ATATATAAAA ATATGAGTAT ATAAATTCAT
GAATATATAG AAATATAAAT AGATCTAATA TATGAATATA TTATATGATG TATATTATGT
ATTATATAGT AATATAATTA TATATTATAC AAAAAGTATA CAAATTAAAT GTATTTTATA
AATTATAAAA TTTATCAATT ATGTATTTTA AATATGTATT TCTGCATAAT GTATATATTA
TATATAATCT ATATTTAAAT TATATATTAT AAATGTATTT TATAAATGTA TACATTTATA
TATTTATATA CTGTAAATGA ATTTTATCAT TTATAATATA TAAATCATAC ATATAAAATG
TTTATATTTC TATAATTTAT AAAATGTTTA ATATATTAAA TATGGTTATT AATGAAATGT
CTAATAATTC AATGTAATAA TTAATTCTAT ATCATTACTT AGTAAGTATA ATACATTATA
TATGTGAATA TAAAGTTGAT GTATATACCG ACAAGAGCCC TTTGCATCTC CCTAGCAATC
CCTGACTCTC TCCCAGCCTC ATGTTTGTAT CTTTCTCCTC AACATGCCCT GTCTCTCTTC

FIGURE 5 (cont'd)

CTACCATTCT ATCCAACTCT CCCGTAACTC TTCCCATCCC TGTTCCTGCT TTTCCCATCT
TTAATTCTCT ATTTCTGACC ATCTCCCTAT TCCAACTCCC TCTCTCCAAC TTTCTCTCCC
CACCGCTGGC TCCACCACTC TCCTTATCAA CCTTCCATTC TCTTGTCCCT TCCCTCCTTG
TCCTTCCCTC CACTTTTCTC CTCATCTCTC CCTTCGCCTC TCTCCCATGT CCCTCCATAT
TTCTGTCACT TCCGTTGCTT TACCCAGATA GGTGCTCATC TCTTCTCCCA TCTTTCTCTT
CCCATCTCAA TTTTCTATCT ACTCTTTACC CATTCAACTC GCCTATTTCA CCTTCATCCC
ATATCCTATC CAGGTCGGAT ACCTTAGACC TTCTCTTTCT TCTCCCCAGT GAATTACCCC
AAAACTCTAC AATGTGCCAA CATCCAACTT CGCTCAGATG AGGAGTGTCG TCAAGTCTAC
CCAGGAAAGA TCACTGACAA CATGTTGTGT GCCGGCACAA AAGAGGGTGG CAAAGACTCC
TGTGAGGTGA GGCCGGGAGG CTGGTGGGTG CCTTGGACAG GATAGAAAGC CAGAATGGAA
GTGACAGATG CTGGGGAAAA AGCTTTGTTT CCAGCCTTAG GGGAACCAAT CTTTATAAGA
TACAATGTCC CCTCACATAG GAGGTCAAGA CAAAAAGGGG TACCCAGGGA TGGCAGGAAT
AATTCATCAT AAGCCCCAGC TTTGACTGAG TGGCTGCCAA GATCCCTGTG TTGAGATGCA
TAAAGGTTGG TATTCTTTCA CTTGTGAGTG ATAGACAACC AACTCAAACT GGCTTAAACA
AAATGCAGGC TTTTGTAACT GAAAATCCAG GTTGTCTGGC TTTAGGCACA GATGGATCCA
GGTATGCAAA TTGTGTGTTT GGAATTCTGT CTTTCTTTTA ACTCTCAGCT CTTCTTTATT
CTGTTTTGGC TTCATTCTCG GTTAGATTCT TCCCATGACA AGATGGCCCC AGCAGCTTTG
AGCTTACATC CTACCCTCTA GGCAACCCTA TTAGAAAGAG AACCTCTCTT TTCCAATAGT
TCACACAAAA GTCTTAAGCA TGATTCTCAC TAGGCTGACC TAAGTCATGT GTCTTGAGCC
ATCACTCCAC CAGAGCTGTG GGATTCTCTG ATGGGCCAAG CCTGAGTCAC ATAGTTAACT
GTGGGTGCTG GAGAGGGGCA GGGACAAACT GCATGGATTG GAAGTGGAGA AGGGCAGTTC
CCCAAATGAA AAAATCAGGA GAGGCTGTTA CCAAAATAAG GGGAAATGGC CAAGTACAGT
AGTTCATGCC TGTAATCCCA GCACTTTGGG AGGCTGAGGT GAGAGGATTA CTTGAGCCCA
GGAGTTTGAG ACCAGCCTGG GCAACATAGT GAGACTCTGT CTCTACAAAA AGAAAAAAAA
GTTTTTAAAT TAGCCAGGTG TGGTGGAGTA CAACTGCAGT CCTAGTTACT CGGGAGGCTG
AGGCAGAAGG ACTATTTGAA CCCAGGAGTT CAAGGCTGCA GTGAGGTATG ATCATGCCAC
TGCACTCCAG CCTGGGTGAT AGAGCAAGGC CCTGTCTCTA AAACAAAAAG AAATAAATAG
AGCAAGACAC TGTCTCTAAT AAATAAATAA ATAAAAATTT AAAAATGAAT GTTTAATTTT
TTAAAAATAA GAGGAAATGG ATACTACATG AGCAAAAAAT AGCCTTCATC AATAAAGAAG
TTGAGATTGG ATTCAGTGAG AAAGAGTATG ATACTATATT AATGATATGT GCCTTGATCG
ATTAGTGATG TCTGCCTTGG GCCCAGGAAG AGAAATAGAC TTACACGTGT GTTGCATACC
CTGCCCAGAT ATGAATGGGT TCACTCAATA GTGAGAGACA CAAATGAGCC TTAAATAGGA
GCAGGGTCAG CTGGTGTGGG GCAGGGGGTG ATTTAGTACC AGGGAAACAA AAATGGGTAT
GAAGTAAGTT GTTACCATTT TAATGAAACT GAGGAACAGA GAAAAACACA GAAATTTCTC
TGTGTCTCTC TTTCTCTGGG CCTATCTCTG TCTTTCTGTC CCTATTTCTG TCTCTTGCTG
TCTGTCCCTC TGTGTTTGTC TTCTTGTCTG TTTCTCACTG TCTTCATTGC TTTCTCTCAC
ACTGTGTGTG TCTGACTCTG CCTCTCTGAG TCTCCTTCTC TGTGTGTGTC TCTCTCCATC
TTTCACTCTC TCCCCACACC TCCCTGTCCC TGCCTTGTTT AGCCCCAGCA AGGACCCACC
TCTCTCTCTC TTTCTTTCCC CAACTCAGGG TGACTCTGGG GGCCCCCTGG TCTGTAACAG
AACACTGTAT GGCATCGTCT CCTGGGGAGA CTTCCCATGT GGGCAACCTG ACCGGCCTGG
TGTCTACACC CGTGTCTCAA GATACGTCCT GTGGATCCGT GAAACAATCC GAAAATATGA
AACCCAGCAG CAAAAATGGT TGAAGGGCCC ACAA*TAA*

FIGURE 6

KLK-L 5

GTCATATTACATGAGGGCTCTGCTAGACTCCGAAAAACAAAAAACAGCAC
AAAGTTCCCTTGTCCTGTGACTCATTCTCTCTCTCTTTCTACCATTTC
TCCTTCCCTGTGTCTTTTTTTTTTCTCTCTGTGGGTTTTATTTAAGCAAT
AGAAGTTCTTAGCAAAGAAAAACTTTATGGAATTAGATTGATCCAGTTCA
TATGTACATATATGAACTCAGTTCAGAAACTCTCTTCTACCCCTGCCTGA
TCACCTATTTGGAAGTCTGTTCCTTCAACTCTTCTTCTCTTTCTGGGACT
CTTTCTAGCTTGGGCTTCCTGCCCCTCCCGTCCACTCTCCTGCTTTCACA
GCCTCTCCTTCCCCCTGCCCCTCCCCTGCACTGCATGGGGATGGGCCCCA
GGTGTCCAAGGTCTCCCCACCCTCCTTTGTCACTGGAGTCAGGATTAGAA
CCCAGCTCCCTAGTCACCTTGAGTCATCAGTCCTGGGGCTGCTGACGGGC
TTGCAGAGGAGAGAGGGAGTGGGGCTGGGTCTTCCCACCCTGGGTCCTTT
CCTCCTTCCCCACTCCGTTTAGCTGTAAAGCTCAATTAAGTGTGATTAGC
TGAGAAGAGTTTCTGCAGAATTAGAGCACGCCCCACCCCTGTCTTCGTGG
TCCCCTTCCCTTAACCCGGAAACTGGATGGGCCAGGACAAAGAGAGTTAA
GAGCTTTGTCAGTGGTCTGTCTGGAGCGACAGATGGAAGGAAAGGGACCG
GTTGAGCAACATGACAGGTGGCTGAGGAGCCAGGTGCAGAGTGGTAGAGT
TGGCTGGCGGAGTGGCCAGCACATGAGAAGACAGGCAGGTAGGTGGACGG
AGAGATAGCAGCGACGAGGACAGGCCAAACAGTGACAGCCACGTAGAGGA
TCTGGCAGACAAAGAGACAAGGTGAGAAGGAGGTAGGCGACTGCCAATGA
GGGAGTGACACACAGGGGAGCAGGTAGAGAGAGGACAAGCAGGTCATCCC
CTTGGTGACCTTCAAAGAGAAGCAGAGAGGGCAGAGGTGGGGGGCACAGG
GAAAGGGTGACCTCTGAGATTCCCCTTTTCCCCCAGACTTTGGAAGTGAC
CCACCATGGGGCTCAGCATCTTTTTGCTCCTGTGTGTTCTTGGTGAGTTC
TCCCGGAGCAGGGAGAGGGCAGGACTGCGACTGGATCCCTTCACCCCCAT
GAGGAGGCCCCACCACCCTCCCCATCTCAGCTCTGGCCCCCAGCCTGGTG
GTGAGGAGGAGAGGGGCTTTCTCTGTGCCTCCATTTACCTGCAGCTCTCA
GGGTACTGCTCACCTCGGTCTCCCCTATTTTTTGATCCCTCTTCCCTTCT
GTCCCTCTCTGAATCTCTGTCTCTCCATTTCCCTCCTATGTGTAAGCATC
TTTCTCCCTGGGTGTCTTTGATGTTTCATGGTCTTTTTCTATCACTGGGT
CTCTCTCTCTTTCTCTCTCTTTCTCGTCTCTCTTTCTCCTCTCTCTCTCC
TGCCTGTTTCTCTCTCTCACTCTGTGTGTCTCTCCATCTCTGTATCTTTT
CTTCCTCTCTCTGACCCATGCCCCTGTCTGTCTCCAGGGCTCAGCCAGGC 1588-1747
AGCCACACCGAAGATTTTCAATGGCACTGAGTGTGGGCGTAACTCACAGC (I)
CGTGGCAGGTGGGGCTGTTTGAGGGCACCAGCCTGCGCTGCGGGGGTGTC
CTTATTGACCACAGGTGGGTCCTCACAGCGGCTCACTGCAGCGGCAGGTA
AGTCCCTTCCTGGGGTGGGCGAAGGGAGGACTATGGGAAGGCAAGCGCTG
GGGGTAGGATCACAAGGGAGGGTGGTGCCCACTGGGAAGAAGCTGATCCT
GCAACAAGAGAGTCTGAGGTTAGACCAGGAGTGGAACTTCCTTAGCAGTG
GGCCTGGGGTGGTGCTGGGCAGGGTGAGGTATGTTGGGTGGAGGGCCGGG
GAGGGTCCTGGAACCTGCCCTCCTGCCTCTCCCATTCCTGCATGTACCCT
TTCTTTCCTATATGACATCTGCCAGTCACCCCAGCCATTCCTTGACCCAG
TCTGGGCCCGGGGCCCAGGTCTCACCCAAGCTCTTTTCTTTTCTTTTT
TTTATTTTTTTGAGACAGGGTCTCGCTCTGTCGCCCAGGCTGCTGTGCAA
TGGCGTGATCACAGCTCACTGCTGTCTCTGCCTCCCAGGTTCAAGTGATT

FIGURE 6 (cont'd)

CTCCTGCCCCAGCCTCCTGAGTAGCTGGGATTACAGGCACCCGCCACCAT
GCCCAGCTAATTTTTGTATTTTTTGTAGAGACAGGGTTTTGCCATGTTGG
CCAGGCTGGTCTCGAACTCCTGGCCTCAAATGACCTGCCCGTCTTGGCCT
CCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGCACCCGGCCAACATGA
CCCAAACTCTTTGTGCAACTTCAGAATCTATGCCTGGCACCTCTCTGGGC
CTCAGTAGACTGATGTTCTGGAATTTTTTTCTTTTTCTTTCTTTTTTTTT
TTTTTTGGAGACAGAGTCTTGCTCTTTCTGTCATCCAAGCTGGAGTGCAG
TGATGCTATCTTGGCTCACTACAGCCTCAACCACCTGGGCTCAAGTGATC
CTCACACCTCAGCCTCCCAAGGAGCTAAGACTACAGGCCTGCGCCACCAC
ACCTGGCTAATTTTTAAATTTTTTTTGTAGAGACAGGGTTTTGCTATGTT
ACCCAGGCTGGTCTCAAACTCCTCAGCTCAAGCAATCTTCCTGCCTTGAC
CTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGTGCCTGGCCTGGAA
CTTTTTTTGTGAAAGGGGAGATCAGATGCAAAGAAACAGAGACTCAGGGA
GAGAGAGGGCCAGCAGCAGGATGCAGAGAGGCCATTCATCAACCCACTCG
TTCAATCATGAACCCACTCGTCCACGCATGAGCATGGAGGGCACATGCTC
CGTGCCAGGCGGTGGGAATAAGGCAGTGAACAAGGTCCACTGATGTCCCT
GCCTTCATGGGCTTCACCAGCCGAGAGAATCAGAAAGAGAGGCCTGGCGC
GGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGA
TCACTTGAGGTCAGGAGTTTGAGACCAGCCTGACACACATGGTGAAACCT
TATCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCATGCTTCTG
TAATCCCAGCTACTTGGGAGGCTGAGGCAGGTGAATTGCTTGAACCTGGG
AGGTGGAGGTTGTAGTGAGCCAAGATGGTGCCACTGCACTCCAGCCTGGG
CGACAGAGCGAGACTCGGTCTTGAAAAAAAAAAAAAAAAAAAAAAGGAGA
GAGAGAGACACAGATGCAGGGACATGGTAGGAGAAACAGGGAACACCCAA
GATGGAAAGAGGGTGATGGAGGTTGGGAATAAGAGCCTGTAAGAGAGACT
CGGAGAATGAGAGTTGCGGGTGAGAGGACAGACAGTGAGGGGCAGAACAG
TGGGGAGCGGCAGGAGCGCCTGAGTGTCCGTGGAGGGGTGCAAGGTGGGG
GACTGCGTGCCTGCCACCCGCTCAGCCGTCGCCACCGGCAG<u>CAGGTACTG</u> 3592-3851
<u>GGTGCGCCTGGGGGAACACAGCCTCAGCCAGCTCGACTGGACCGAGCAGA</u> (2)
<u>TCCGGCACAGCGGCTTCTCTGTGACCCATCCCGGCTACCTGGGAGCCTCG</u>
<u>ACGAGCCACGAGCACGACCTCCGGCTGCTGCGGCTGCGCCTGCCCGTCCG</u>
<u>CGTAACCAGCAGCGTTCAACCCCTGCCCCTGCCCAATGACTGTGCAACCG</u>
<u>CTGGCACCGAGTGCCACGTCTCAGGCTGGGGCATCACCAACCACCCACGG</u>
AGTAAGGGGCCCAGGGCCAGGGGTCAGGGGTCAGGATGGGTACAAGTCTG
GGATGCAGGGCGAGAGGTCGAATCATGACACCTCAGAGGAAGGATGGGTA
AAGGGTCAGGGTGTGGGATGGGACATCAGGATCATGGTTTGGGGTCAGAG
ATTATGGTGGATTGGGGTCTTGGGAGCCAAAGGGGTTAAAGGACTGGGTA
TGAAGTCAGGGATCAGAGGTCAGAGGTCAGAGTGTGTCAGAGGTCATCAC
ACTGGAGCAAAAGGCATATATATATATATATGTATGTATAGGATATGGGC
ATTGTGGGTCATGGGTCTGGGGTTAGAGGTCACCGTAGAATTAAGGTCAT
GGGATCCAGAGGTTGTACAATCTGGTCAAAATCTGAGGATGGAAATTGGG
ATTCTATCCAAAATCACATATCTGAGATTGGAGGTCATAGCGTTTGGGGT
GTGGGGCCCGAAGTTTGGGGTCATGGAGGCTGGGGCCCAATAAACTAGGA
TCAGGGGACACTGGCGTTGGAAGCAGTGAGGTTTGGAAGATGCAGAGCTG
AGGTTGGAGGTTAAGGTAAAGACAGGGACATGGGGTCAGGAGACAGAAGA
TATGAGATCAAGCTGGGATCATAAGGTAATAAGACAGAAGGTCAAAGATC

FIGURE 6 (cont'd)

ACAGTAGCTGGCATTGAAGAGGGTCAGGTCTGGATTCGTTGTCTCTGACG
CTGGAGAGACAAGAAAGTTCTTGAGTTATGCCACTCAAAGTCAAATGTCA
AAGATCAAAGAGACCGTCAATCATCTGGGGTCATGATTCATATGAAATTA
AGTCATAAATATGTAACTTGGAGGTTTCGGGATTGTAGTACAGGTCGGTG
AGGGGCAGGGGTATTGACATGGATGGGCCACATCCAGGGAAGAGGGACGT
GGCCTCAAAGTGGGGAGATTTAGGGGACCCTGCAGCACGCATGTTCTCTC
TCCAG**ACCCATTCCCGGATCTGCTCCAGTGCCTCAACCTCTCCATCGTCT** 4806-4939
CCCATGCCACCTGCCATGGTGTGTATCCCGGGAGAATCACGAGCAACATG (3)
GTGTGTGCAGGCGGCGTCCCGGGGCAGGATGCCTGCCAGGTGAGCCAGTG

FIGURE 7

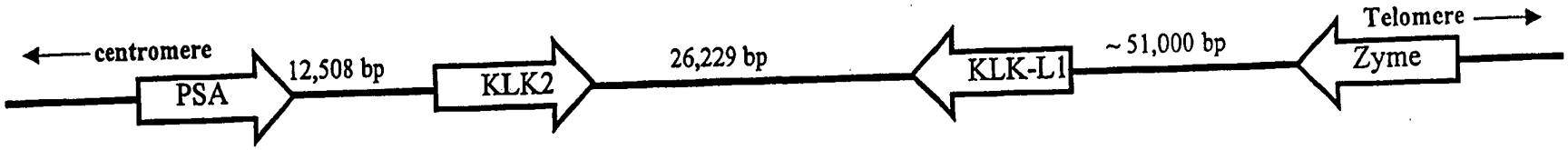


FIGURE 8

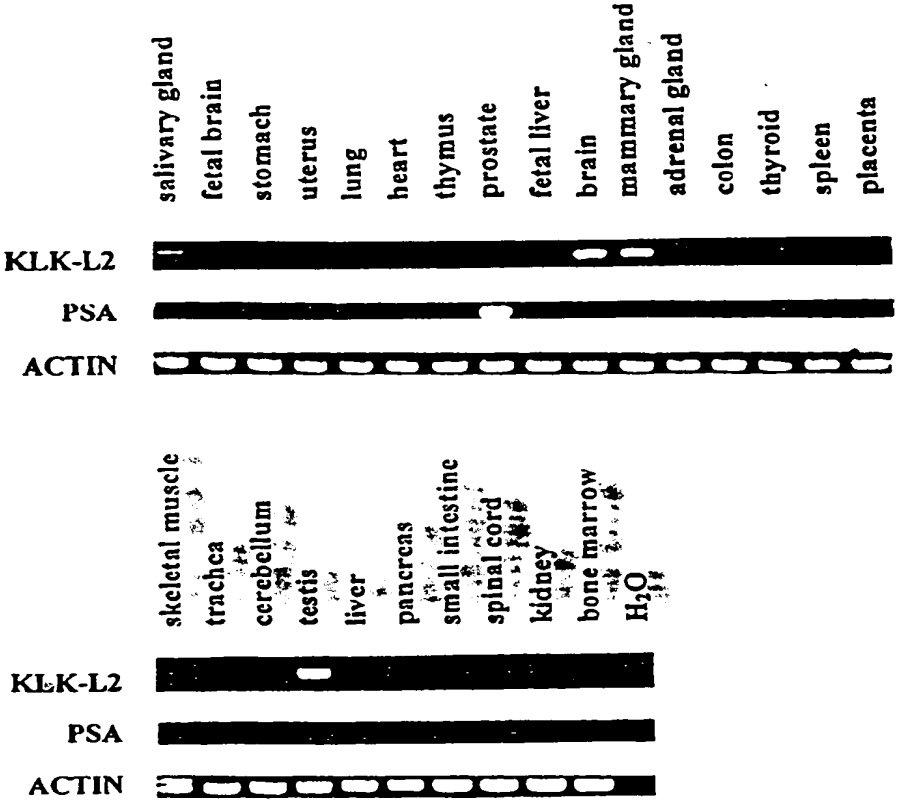

salivary gland
fetal brain
stomach
uterus
lung
heart
thymus
prostate
fetal liver
brain
mammary gland
adrenal gland
colon
thyroid
spleen
placenta
KLK-L2
PSA
ACTIN
skeletal muscle
trachea
cerebellum
testis
liver
pancreas
small intestine
spinal cord
kidney
bone marrow
H2O
KLK-L2
PSA
ACTIN

FIGURE 9

TGACCCGCTG TACCACCCCA GCATGTTCTG CGCCGGCGGA GGGCAAGACC AGAAGGACTC
CTGCAACGGT GACTCTGGGG GGCCCCTGAT CTGCAACGGG TACTTGCAGG GCCTTGTGTC
TTTCGGAAAA GCCCCGTGTG GCCAAGTTGG CGTGCCAGGT GCCTACACCA ACCTCTGCAA
ATTCACTGAG TGGATAGAGA AAACCGTCCA GGCCAGTTAA CTCTGGGGAC TGGGAACCCA
TGAAATTGAC CCCCAAATAC ATCCTGCGGA AGGAATTC

FIGURE 10

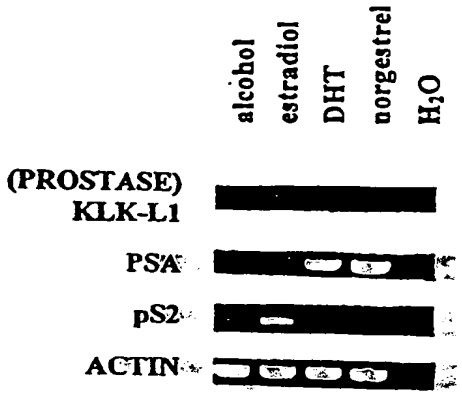

alcohol
estradiol
DHT
norgestrel
H_2O
(PROSTASE)
KLK-L1
PSA
pS2
ACTIN

FIGURE 11

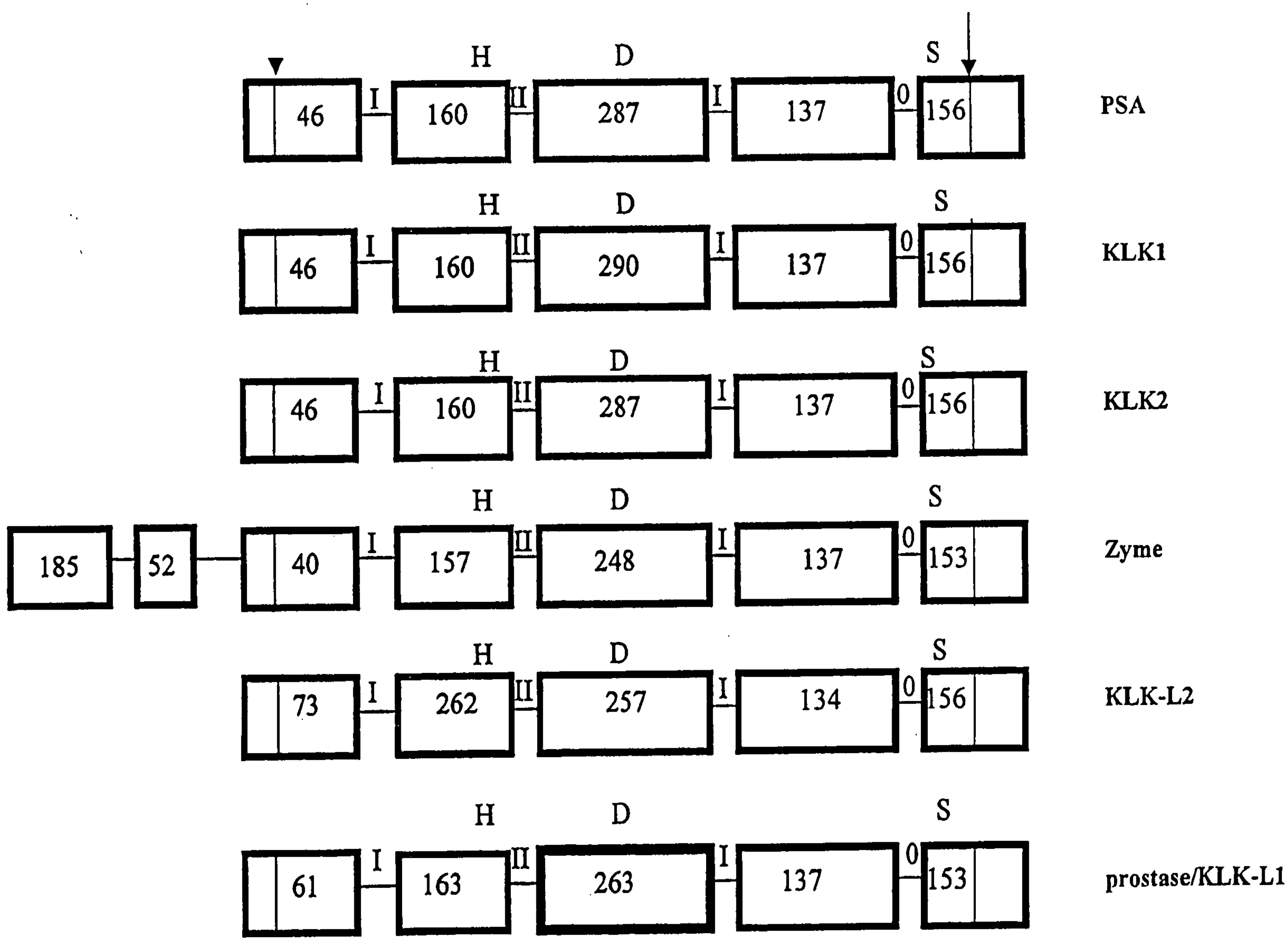


FIGURE 12

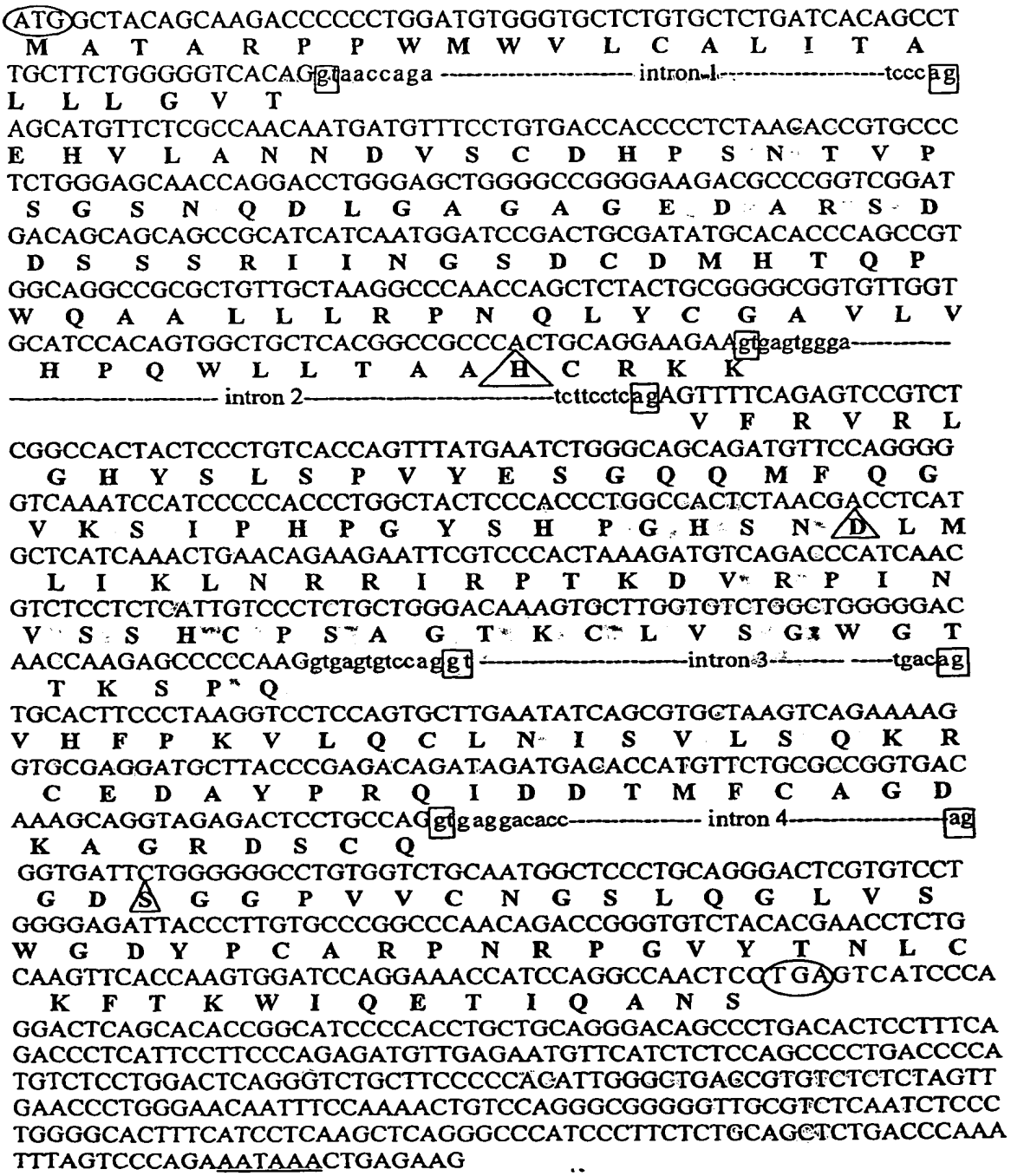

ATG GCTACAGCAAGACCCCCCTGGATGTGGGTGCTCTGTGCTCTGATCACAGCCT
M A T A R P P W M W V L C A L I T A
TGCTTCTGGGGGTCACAGgtaaccaga ---------- intron 1 ----------tcccag
L L L G V T
AGCATGTTCTCGCCAACAATGATGTTTCCTGTGACCACCCCTCTAAGACCGTGCCC
E H V L A N N D V S C D H P S N T V P
TCTGGGAGCAACCAGGACCTGGGAGCTGGGGCCGGGGAAGACGCCCGGTCGGAT
S G S N Q D L G A G A G E D A R S D
GACAGCAGCAGCCGCATCATCAATGGATCCGACTGCGATATGCACACCCAGCCGT
D S S S R I I N G S D C D M H T Q P
GGCAGGCCGCGCTGTTGCTAAGGCCCAACCAGCTCTACTGCGGGGCGGTGTTGGT
W Q A A L L L R P N Q L Y C G A V L V
GCATCCACAGTGGCTGCTCACGGCCGCCCACTGCAGGAAGAAgtgagtggga----------
H P Q W L L T A A H C R K K
---------- intron 2----------tcttcctcagAGTTTTCAGAGTCCGTCT
V F R V R L
CGGCCACTACTCCCTGTCACCAGTTTATGAATCTGGGCAGCAGATGTTCCAGGGG
G H Y S L S P V Y E S G Q Q M F Q G
GTCAAATCCATCCCCCACCCTGGCTACTCCCACCCTGGCCACTCTAACGACCTCAT
V K S I P H P G Y S H P G H S N D L M
GCTCATCAAACTGAACAGAAGAATTCGTCCCACTAAAGATGTCAGACCCATCAAC
L I K L N R R I R P T K D V R P I N
GTCTCCTCTCATTGTCCCTCTGCTGGGACAAAGTGCTTGGTGTCTGGCTGGGGGAC
V S S H C P S A G T K C L V S G W G T
AACCAAGAGCCCCCAAGgtgagtgtccaggt ----------intron 3---------- tgacag
T K S P Q
TGCACTTCCCTAAGGTCCTCCAGTGCTTGAATATCAGCGTGCTAAGTCAGAAAAG
V H F P K V L Q C L N I S V L S Q K R
GTGCGAGGATGCTTACCCGAGACAGATAGATGACACCATGTTCTGCGCCGGTGAC
C E D A Y P R Q I D D T M F C A G D
AAAGCAGGTAGAGACTCCTGCCAGgtgaggacacc---------- intron 4----------ag
K A G R D S C Q
GGTGATTCTGGGGGGCCTGTGGTCTGCAATGGCTCCCTGCAGGGACTCGTGTCCT
G D S G G P V V C N G S L Q G L V S
GGGGAGATTACCCTTGTGCCCGGCCCAACAGACCGGGTGTCTACACGAACCTCTG
W G D Y P C A R P N R P G V Y T N L C
CAAGTTCACCAAGTGGATCCAGGAAACCATCCAGGCCAACTCCTGAGTCATCCCA
K F T K W I Q E T I Q A N S
GGACTCAGCACACCGGCATCCCCACCTGCTGCAGGGACAGCCCTGACACTCCTTTCA
GACCCTCATTCCTTCCCAGAGATGTTGAGAATGTTCATCTCTCCAGCCCCTGACCCCA
TGTCTCCTGGACTCAGGGTCTGCTTCCCCCACATTGGGCTGAGCGTGTCTCTCTAGTT
GAACCCTGGGAACAATTTCCAAAACTGTCCAGGGCGGGGGTTGCGTCTCAATCTCCC
TGGGGCACTTTCATCCTCAAGCTCAGGGCCCATCCCTTCTCTGCAGCTCTGACCCAAA
TTTAGTCCCAGAAATAAACTGAGAAG

FIGURE 13

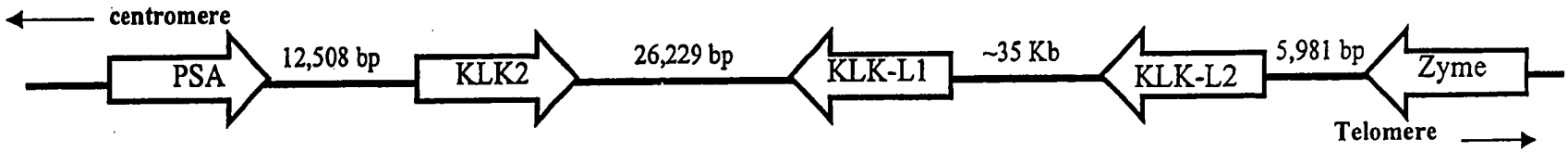


FIGURE 14

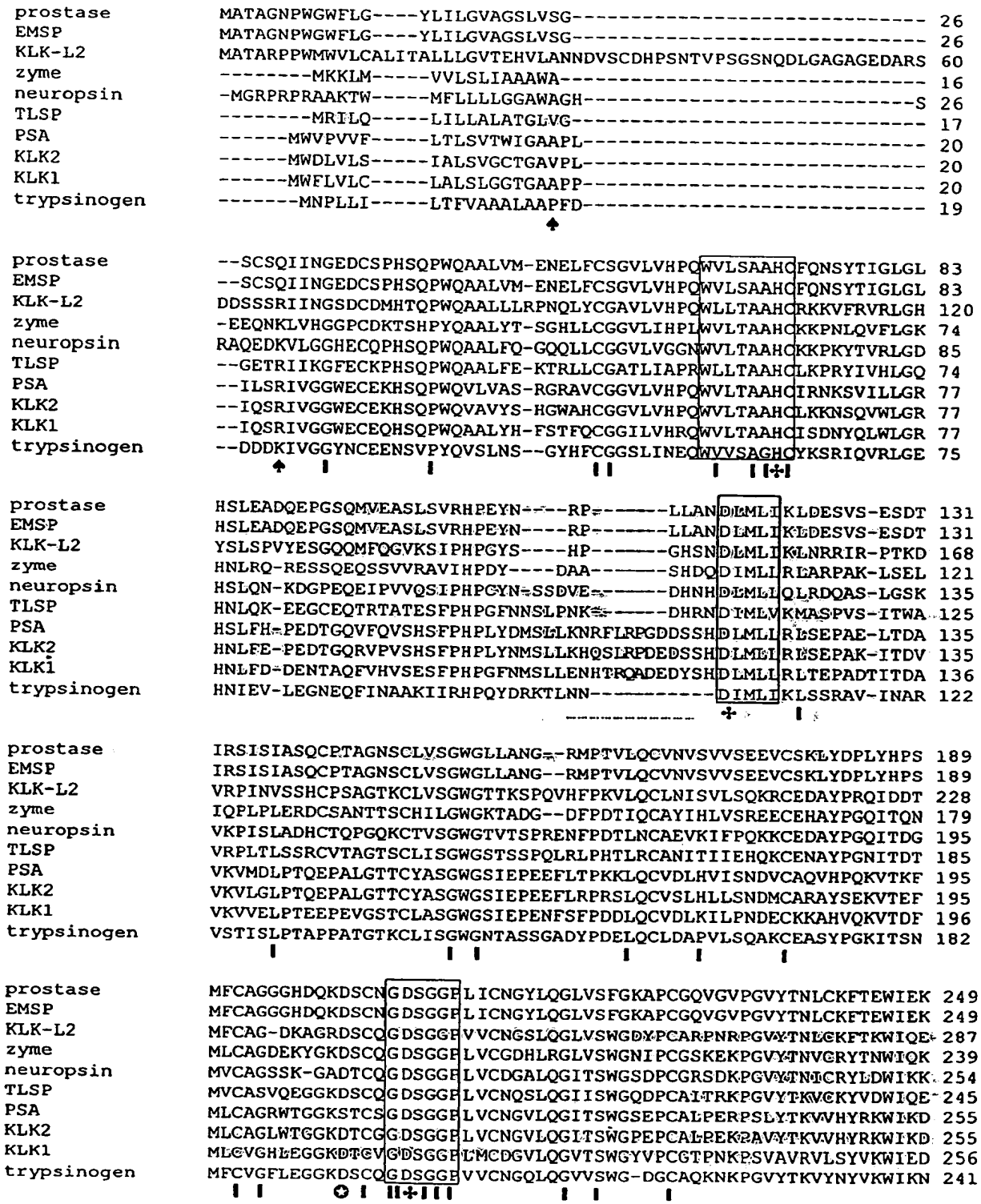
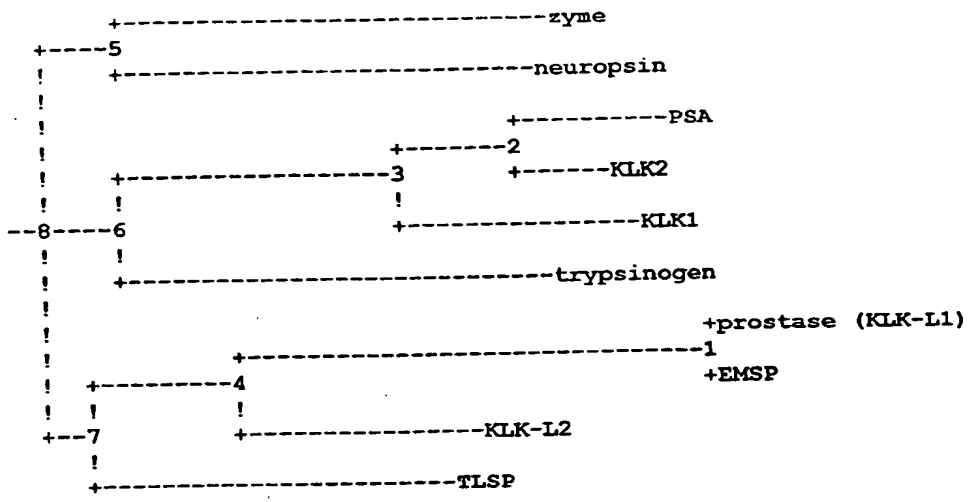

prostase MATAGNPWGWFLG----YLILGVAGSLVSG---------- 26
EMSP MATAGNPWGWFLG----YLILGVAGSLVSG---------- 26
KLK-L2 MATARPPWMWVLCALITALLLGVTEHVLANNDVSCDHPSNTVPSGSNQDLGAGAGEDARS 60
zyme --------MKKLM-----VVLSLIAAAWA---------- 16
neuropsin -MGRPRPRAAKTW-----MFLLLLGGAWAGH----------S 26
TLSP --------MRILQ-----LILLALATGLVG---------- 17
PSA ------MWVPVVF-----LTLSVTWIGAAPL---------- 20
KLK2 ------MWDLVLS-----IALSVGCTGAVPL---------- 20
KLK1 ------MWFLVLC-----LALSLGGTGAAPP---------- 20
trypsinogen -------MNPLLI-----LTFVAAALAAPFD---------- 19
prostase --SCSQIINGEDCSPHSQPWQAALVM-ENELFCSGVLVHPQWVLSAAHCFQNSYTIGLGL 83
EMSP --SCSQIINGEDCSPHSQPWQAALVM-ENELFCSGVLVHPQWVLSAAHCFQNSYTIGLGL 83
KLK-L2 DDSSSRIINGSDCDMHTQPWQAALLLRPNQLYCGAVLVHPQWLLTAAHCRKKVFRVRLGH 120
zyme -EEQNKLVHGGPCDKTSHPYQAALYT-SGHLLCGGVLIHPLWVLTAAHCKKPNLQVFLGK 74
neuropsin RAQEDKVLGGHECQPHSQPWQAALFQ-GQQLLCGGVLVGGNWVLTAAHCKKPKYTVRLGD 85
TLSP --GETRIIKGFECKPHSQPWQAALFE-KTRLLCGATLIAPRWLLTAAHCLKPRYIVHLGQ 74
PSA --ILSRIVGGWECEKHSQPWQVLVAS-RGRAVCGGVLVHPQWVLTAAHCIRNKSVILLGR 77
KLK2 --IQSRIVGGWECEKHSQPWQVAVYS-HGWAHCGGVLVHPQWVLTAAHCLKKNSQVWLGR 77
KLK1 --IQSRIVGGWECEQHSQPWQAALYH-FSTFQCGGILVHRQWVLTAAHCISDNYQLWLGR 77
trypsinogen --DDDKIVGGYNCEENSVPYQVSLNS--GYHFCGGSLINEQWVVSAGHCYKSRIQVRLGE 75
prostase HSLEADQEPGSQMVEASLSVRHPEYN----RP-------LLANDLMLIKLDESVS-ESDT 131
EMSP HSLEADQEPGSQMVEASLSVRHPEYN----RP-------LLANDLMLIKLDESVS-ESDT 131
KLK-L2 YSLSPVYESGQQMFQGVKSIPHPGYS----HP-------GHSNDLMLIKLNRRIR-PTKD 168
zyme HNLRQ-RESSQEQSSVVRAVIHPDY----DAA-------SHDQDIMLLRLARPAK-LSEL 121
neuropsin HSLQN-KDGPEQEIPVVQSIPHPGYN-SSDVE-------DHNHDLMLLQLRDQAS-LGSK 135
TLSP HNLQK-EEGCEQTRTATESFPHPGFNNSLPNK-------DHRNDIMLVKMASPVS-ITWA 125
PSA HSLFH-PEDTGQVFQVSHSFPHPLYDMSLLKNRFLRPGDDSSHDLMLLRLSEPAE-LTDA 135
KLK2 HNLFE-PEDTGQRVPVSHSFPHPLYNMSLLKHQSLRPDEDSSHDLMLLRLSEPAK-ITDV 135
KLK1 HNLFD-DENTAQFVHVSESFPHPGFNMSLLENHTRQADEDYSHDLMLLRLTEPADTITDA 136
trypsinogen HNIEV-LEGNEQFINAAKIIRHPQYDRKTLNN----------DIMLIKLSSRAV-INAR 122
prostase IRSISIASQCPTAGNSCLVSGWGLLANG--RMPTVLQCVNVSVVSEEVCSKLYDPLYHPS 189
EMSP IRSISIASQCPTAGNSCLVSGWGLLANG--RMPTVLQCVNVSVVSEEVCSKLYDPLYHPS 189
KLK-L2 VRPINVSSHCPSAGTKCLVSGWGTTKSPQVHFPKVLQCLNISVLSQKRCEDAYPRQIDDT 228
zyme IQPLPLERDCSANTTSCHILGWGKTADG--DFPDTIQCAYIHLVSREECEHAYPGQITQN 179
neuropsin VKPISLADHCTQPGQKCTVSGWGTVTSPRENFPDTLNCAEVKIFPQKKCEDAYPGQITDG 195
TLSP VRPLTLSSRCVTAGTSCLISGWGSTSSPQLRLPHTLRCANITIIEHQKCENAYPGNITDT 185
PSA VKVMDLPTQEPALGTTCYASGWGSIEPEEFLTPKKLQCVDLHVISNDVCAQVHPQKVTKF 195
KLK2 VKVLGLPTQEPALGTTCYASGWGSIEPEEFLRPRSLQCVSLHLLSNDMCARAYSEKVTEF 195
KLK1 VKVVELPTEEPEVGSTCLASGWGSIEPENFSFPDDLQCVDLKILPNDECKKAHVQKVTDF 196
trypsinogen VSTISLPTAPPATGTKCLISGWGNTASSGADYPDELQCLDAPVLSQAKCEASYPGKITSN 182
prostase MFCAGGGHDQKDSCNGDSGGPLICNGYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEK 249
EMSP MFCAGGGHDQKDSCNGDSGGPLICNGYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEK 249
KLK-L2 MFCAG-DKAGRDSCQGDSGGPVVCNGSLQGLVSWGDYPCARPNRPGVYTNLCKFTKWIQE 287
zyme MLCAGDEKYGKDSCQGDSGGPLVCGDHLRGLVSWGNIPCGSKEKPGVYTNVCRYTNWIQK 239
neuropsin MVCAGSSK-GADTCQGDSGGPLVCDGALQGITSWGSDPCGRSDKPGVYTNICRYLDWIKK 254
TLSP MVCASVQEGGKDSCQGDSGGPLVCNQSLQGIISWGQDPCAITRKPGVYTKVCKYVDWIQE 245
PSA MLCAGRWTGGKSTCSGDSGGPLVCNGVLQGITSWGSEPCALPERPSLYTKVVHYRKWIKD 255
KLK2 MLCAGLWTGGKDTCGGDSGGPLVCNGVLQGITSWGPEPCALPEKPAVYTKVVHYRKWIKD 255
KLK1 MLCVGHLEGGKDTCVGDSGGPLMCDGVLQGVTSWGYVPCGTPNKPSVAVRVLSYVKWIED 256
trypsinogen MFCVGFLEGGKDSCQGDSGGPVVCNGQLQGVVSWG-DGCAQKNKPGVYTKVYNYVKWIKN 241

FIGURE 15

(A)


zyme
5
neuropsin
PSA
2
3
KLK2
KLK1
8
6
trypsinogen
prostase (KLK-L1)
1
EMSP
4
KLK-L2
7
TLSP

(B)

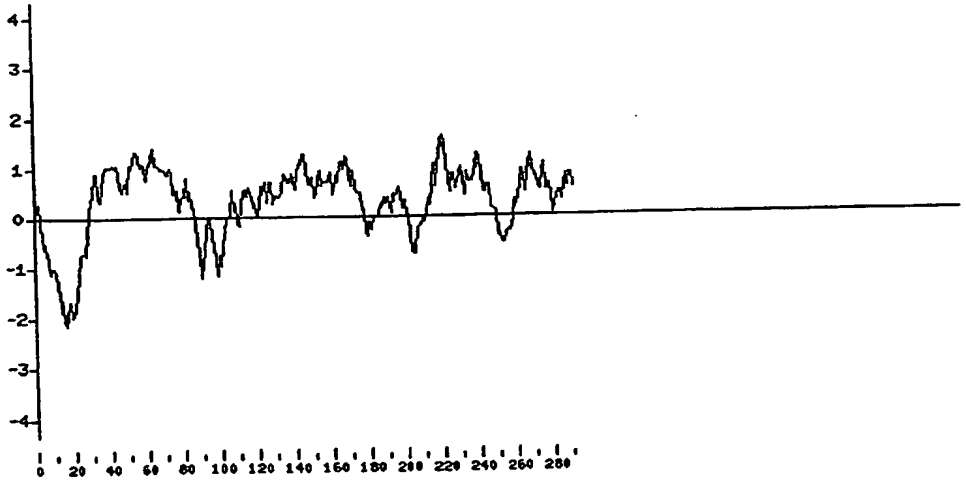

4
3
2
1
0
-1
-2
-3
-4
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280

FIGURE 16

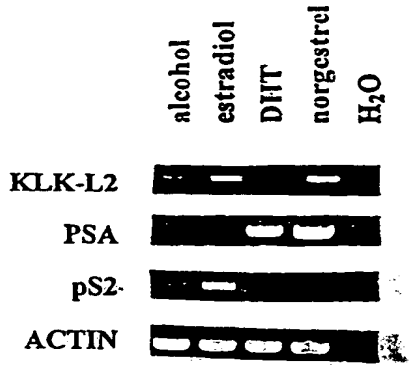


FIGURE 17

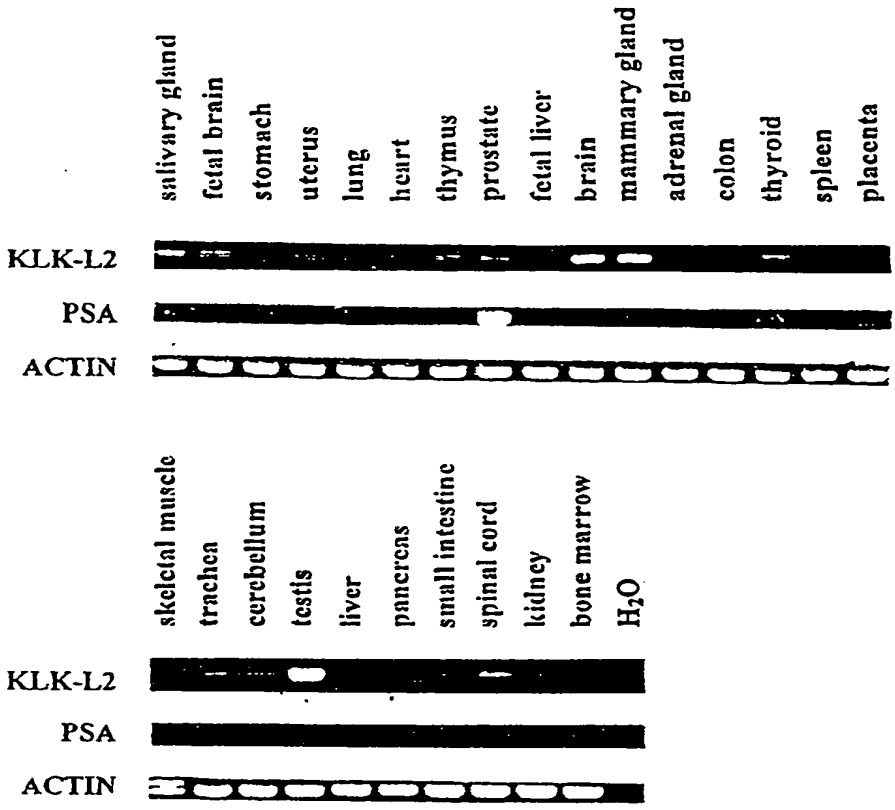

salivary gland
fetal brain
stomach
uterus
lung
heart
thymus
prostate
fetal liver
brain
mammary gland
adrenal gland
colon
thyroid
spleen
placenta
KLK-L2
PSA
ACTIN
skeletal muscle
trachea
cerebellum
testis
liver
pancreas
small intestine
spinal cord
kidney
bone marrow
H2O
KLK-L2
PSA
ACTIN

FIGURE 18

```
FEATURES             Location/Qualifiers
     source          1..8280
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19q13.3-q13.4"
     mRNA            join(3714..3885,5715..5968,6466..6602,7258..7410)
                     /gene="KLK-L6"
     gene            3714..7410
                     /gene="KLK-L6"
                     /note="kallikrein-like serine protease"
     CDS             join(3714..3885,5715..5968,6466..6602,7258..7410)
                     /gene="KLK-L6"
                     /note="serine protease, kallikrein-like"
                     /codon_start=3
                     /product="Kallikrein-like 6"

/translation="MTQSQEDENKIIGGHTCTRSSQPWQAALLAGPRRRFLCGGALLS

GQWVITAAHCGRPILQVALGKHNLRRWEATQQVLRVVRQVTHPNYNSRTHDNDLMLLQ

LQQPARIGRAVRPIEVTQACASPGTSCRVSGWGTISSPIARYPASLQCVNINISPDEV-

CQKAYPRTITPGMVCAGVPQGGKDSCQGDSGGPLVCRGQLQGLVSWGMERCALPGYPG
      VYTNLCKYRSWIEETMRDK"
```

FIGURE 19

```
BASE COUNT     1804 a    2392 c    2246 g    1838 t
ORIGIN
        1 atcgtgtaat caccgccaca tccagtgcaa agctgattcg tcaccacaga gcagctccct
       61 cctgccaccc catccctggg tcccaagaga accctttctt aaaagaggga gttcttgacg
      121 ggtgtggtgg ctcatgcctg taatccttgc actttgggag gccaaggagg gtggatcatt
      181 tgaggtcagg agtttgagac cagactggcc aacatggtga aaccctgtct ttactaaaaa
      241 tacaaaaaaa tgagcggggc atggtggtgg gtgcctatag ccccagctac tcaggaggct
      301 gaggcaggag aatcgcttga acccaggagg cagaggttgc agtgagccga gattgagcca
      361 ctgcactcca gccggggcta aagagtgaga ctctgtctca aaaaaaaaaa aaagaaaaag
      421 aaaaaaagaa aaaaaaataa aataaataaa taaataaaat aaatttaaaa atttaaaaat
      481 aaagaggggg ttcttgtgtt gatgccgagc ctgaaccaag gcagaggagg ccgggaaggc
      541 ttcccaaggc cttcagctca aagcagggag gcccatagtt aaacagaaac agttcaggaa
      601 tcacagaaag gcacctgggg agagatgggt gtgtggctcc agatgcaggt gcccagacag
      661 tgcgtcccca ggtgtacaga cagacccagg ccaagctcca gctcaaagag ccagcctagg
      721 ggggtgccga ggtggaggga ggctgagtca ggctgaggcc ggggaacagt tggggtagcc
      781 aagggaggca agcagcctcc tgagtcacca cgtggtccag gtacggggct gcccaggccc
      841 agagacggac acaagcactg gggaatttaa ggggctaggg gaggggctga ggagggtagg
      901 ccctccccca aatgaggatg gaaccccccc aactccagaa ccccccctgca ggctggccag
      961 aatccttccc catctcattc actctgtctc tcctgctctc tgccgtctcc tattttgaat
     1021 ttccaacccc gtctgttaag actgtccttc tgtctctgaa tctctgtccc cttctctttc
     1081 tgggtctctc tccctctccc tctgggtctc tgtccccctc tctgggtctc tgtcactctc
     1141 tctttgcatc tccagctctc actttgtctc tgcacctagc agatcccaag ctggggaatg
     1201 ccagttctgg caccaacctt cctgctccct gctggggcct ctgctccccc atctctcagg
     1261 agtcgaaagt gagaaagcaa ggtgggcagc tctgctccag gtccaggtat ctcccgccca
     1321 cctcctgccc gtcctctatc ccacccctcc tctccatctc tccctggcgc tgccatctct
     1381 catctaggcc tccgtctcct ctgtcattgt ccccatcccc tgtaggtgcc catccttccc
     1441 gtctcccctc tgccatcggc ctgcctgtcc catcctcttt ctcccaccat gtcccgttct
     1501 cttccacgtc tcatgcccgc actgccttca tcatcatcgc tgttgttctg tgtgtgtttg
     1561 tggtgagtgc cgcatggtgg gggcgtctcg gcctctctcc tctctctcca ctgttttctc
     1621 tttctgtgtg tctgtttcca ttctatctcc accttcttcc ctccgtcttt tgcttttcta
     1681 tctccacttc tccacacccc tctctccctg cgtctctgtg tctccctctt cctctgtctt
     1741 gtttttttcc caccgtctgc ctcttctgtt ccctgtcaca tccaacttcc accggtttct
     1801 ccagctctct cctcagttcc ttctctcatg agcacacctg cctctgtgct cgtattcctg
     1861 gactcctctc tctccactgt catatcttct cattcatttt cccagtctct ctctgtctct
     1921 tgctctcccc ctctctgtca ctctgtctct gtctctctct ttctctctct ctctctgtgt
     1981 ctctctgtct ggctctctct ctgtctctct ctccatctct ctctctctct cccccccgtc
     2041 accctgtctc tgtctctctc tgtctgtgtg tctctctgtc tttctctctc tccatctctc
     2101 tctgtctctc tctctctctc tctctctctc cctctctccc tcctcccgtg actccctctc
     2161 tcagtccatc tcttcctccc tctctcagcc ccttcgtgcc ctttcctctg acactcccca
     2221 ccctggtttc ctgactccac cactagatcc accacctcca gcaactggga accctcccct
     2281 gcccaccctg ccctggggtc ccctcccagg attccttcta gattatagca tcttccctgg
     2341 gcgggttctc atgaacaatt gtggctgctt ttttggccag acaggggagg gaggggatgg
     2401 gatcagggag tcctggaatg ggaactaggc aataaaaaaa aaaaaatgtc agaagcaggg
     2461 cggcgggagg tgggggcagg gccagctgtc cttaccaggg ataaaaggct ttgccagtgt
     2521 gactaggaag agagacacct cccctccttc cttcatcaag acatcaagga gggacctgtg
     2581 ccctgctcca catcctccca cctgccgccc gcagagcctg caggccccgc ccccctcgtc
     2641 tctggtccct acctctctgc tgtgtcttca tgtccctgag ggtcttgggc tctgggtaag
     2701 tgcccccttgc tgtctctgcc tctcagcccc cggttctgtt gaaggttcct tctctctcac
     2761 tttttctctg catttgacag gacctggccc tcagccccta aaatgttcct cctgctgaca
     2821 gcacttcaag tcctggctat aggtaagaga acggttgggt atgacacaag ggggtcccct
     2881 ggagactctg agaagagatg gggatgggtc cttggggccc ctggatgctc atggtgacct
     2941 cataagaaag agcagggagt ggtttggggg tcatggtggg ggaacgtgct ggaggcctaa
```

FIGURE 19 (cont'd)

```
3001 attcctagtt gtggaggtgc tagggaattg tggggccggg gagagaggtg tttataaggt
3061 ctggtgcaaa atacataagg aatcttaggg aactattagg tcctgagtgg gtcatagcag
3121 aaagatcacg gggctctacc tgactgtgtt aggaaagaaa caatgtcaga aagatgtttt
3181 gttgtcagag ggaaggtgga gaaggatgat gggatggcgg gatcgtggca tggggtggcg
3241 ggatcgtggc atgggtgtgt gaggtggatg ggggcaagtg tggggcaaga gatggcggat
3301 ccttggggtc ccactgagtg ggaacgttgg ggaggagaca gggaggtcct tgaatgtgtt
3361 ggggaaggac tcattggggg gaaatgtggc atatttcgag aagtgatcac agaaattatg
3421 ggagcataga gctaagggtc gtagatgtag caaggccctg gataaggtgg ccacggcaca
3481 aaataagaga tgctacggag gtgacttggg aggtgagtca gaaagctctc cgtgctgggg
3541 caataacggg gtcaatattg ggcatgtctc accctgggtg ggacagatag aggcgggcag
3601 tttaggggtt agaccaaaag gaaggggatt tgtcagtttt ggaatcctac aaacttgtgg
3661 agtggagagt gtttgctcat ctactttccc cacccaatcc tgtccactcc tagccatgac
3721 acagagccaa gaggatgaga acaagataat tggtggccat acgtgcaccc ggagctccca
3781 gccgtggcag gcggccctgc tggcgggtcc caggcgccgc ttcctctgcg gaggcgccct
3841 gctttcaggc cagtgggtca tcactgctgc tcactgcggc cgcccgtaag tgacccctc
3901 ccctgtccct gtacctagtg aattccagag tctaaagccc tagagctgag ctgagaacct
3961 ggatctctgt atagaaccca atgtagtggc tggctcctgg tttgaggtct agagaagagc
4021 ctggaacaaa aacacagctc gggatgtggg ctcctccata aatctcgaac tcagcatagg
4081 ttctgaaagc agatgggcag cttggaaccc atggacctgc tgagaaccga acatctgatc
4141 cagtgattct tccagaggcc acacattaca tcgagaccaa gcttagccca ttccagattg
4201 gtggctgaat tcaggacccc gtctacattc agaaactcag gacactacgt agaactcaga
4261 gcccagttca ggacctgcag tctagccata aatccagaac tagaacgctg ctcacagctg
4321 gaacatacaa ctctaagaat agaggcaaaa cctggaggct gtttcacacc caaggtttag
4381 ttcagagtct agtctatagc tccgctatga gcagacttca acccagtgtt tgaatcccag
4441 aatgtggcgg gtgcggtggc tcatgcctat aatcctagca ctttgggatg ctgaggcagg
4501 cagatcacct gaggtcagga gttcgagacc agcctgagca acatagagaa accctgtctc
4561 tactaaaaat gcaaaattag ccaggcatgg tggcacatgc ctgtaatccc agccactcgg
4621 gaggctgagg caggagaatc acttgaacct gggaggcgga ggttgcagtg agtcaagatc
4681 gcaccattgc actccaggct aggcaacaag agcgaaactc catatcaatc aatcaatcaa
4741 taaatcccag aatgcagatc ctaatcagaa gccccatata aaacctagac ccctcctaaa
4801 ttctagatct gaacttacaa cccagacccc agccaagagg tcaaaatgcc tataagccat
4861 atctatgcca taaacaggtc agtctagaac ctagagatca aagctcaggc cagagtctag
4921 aatataaagg ccagaatgca aaccagactc tagaatcttg gatccgggcc ataacctaga
4981 gctccaacta gaacccagag cccaacctga ggtcaagggc tagggccaga gtccagaacc
5041 aagagcccta taatccaata tgaaacagac ctgtagaggc tgggtgcggt ggctcacgcc
5101 tgtaatccca gcactttggg aggctgaggc gggagaatca cttgaactgg gagttggagg
5161 tcgagagtga gctgagatcg tgccactgca ctccagccta ggtgacagag cgagactcca
5221 tcacaaaaaa aaaataaata aataaatcaa gtcataatcc aggttcgatc tagaatcctg
5281 atcttagcat agagtcaaaa gtttaagatg tctagaactc agaacccagg ctagaaacag
5341 aatggtgcct actccggaat atcagttccg atttagagcc tagactcata acgcagtttc
5401 gcttaggact caatgcaccg agcccagcac agaccctggc acggagccaa gctctcccaa
5461 tcatcacctt cttcccaagc caggagctgg agcccagccc aagagcggaa ggagaggcag
5521 ctggggctgg gccgagagaa tgccctggcc atggggaagg gcacaggagg ccaagaatgc
5581 tcggcctgca gttagtgaga agcaggctag acctcgggga agactcgtca cccggccagg
5641 gaaccgggct ggagggtggg gaggagtctc tggctcagac cctgagcagc gcttctcttg
5701 ggggtcgtgg ccaggatcct tcaggttgcc ctgggcaagc acaacctgag gaggtgggag
5761 gccacccagc aggtgctgcg cgtggttcgt caggtgacgc accccaacta caactcccgg
5821 acccacgaca acgacctcat gctgctgcag ctacagcagc ccgcacggat cgggagggca
5881 gtcaggccca ttgaggtcac ccaggcctgt gccagccccg ggacctcctg ccgagtgtca
5941 ggctggggaa ctatatccag ccccatcggt gaggactcct gcgtcttgga aagcagggga
6001 ctgggcctgg gctcctgggt ctccaggagg tggagctggg gggactgggg ctcctgggtc
6061 tgagggagga ggggctgggc ctggactcct gggtctgagg gaggaggggg ctgaggcctg
6121 gactcctggg tctcaaggag gaggagctgg gcctggactc atacgtctga gggaggaggg
6181 gctggagcct ggactcctgg gtctcaagga ggaggggctg ggcctggact tctgggtctg
6241 agggaggagg ggctggggac ctggactccc gggtctgagg gaggagggac tgggggtctg
6301 gactcctggg tctgagggag gaggggctgg gggcctggac tcctgggtct gagggaggag
6361 gtgctggggc tggactcctg ggtcggaagg aggaggggct gggggcctgg acccttgggt
```

FIGURE 19 (cont'd)

```
6421 cttatgggag ggtagaccca gttataaccc tgcagtgtcc cccagccagg taccccgcct
6481 ctctgcaatg cgtgaacatc aacatctccc cggatgaggt gtgccagaag gcctatccta
6541 gaaccatcac gcctggcatg gtctgtgcag gagttcccca gggcgggaag gactcttgtc
6601 aggtaaggcc caggatggga gctgtggtag ggattatttg ggactgggat ttaagcaaat
6661 gatgtcagga gcatggaagt ctgcagaggt cttcagaaga gagtgaaccg caggcacaga
6721 gagattccga tagccaggcc accctgcttc ctagccctgt gccccctggg taatggactc
6781 agagcattca tgcctcagtt tcctcatctg tcaggtggga gtaaccctct tagggtagtt
6841 ggtggaatgg gatgaggcag gttggggaaa gatcgcagag tggcctctgc tcatatgggt
6901 ctgggaaagg ctgtgctgag gcttctagaa atcttaatgc atccttgagg gaggcagaga
6961 tggggaaata gaaaaagaga gacacacaaa tgttctacag ttggagcgaa cagagagggg
7021 cctggtgaga ttcaagggac aggcaggtgc acacagagac agagccagac ccagcggaga
7081 gggaaggaag tgccccgacc tccggggctg agacctcaga gctggggcag gactgtgtcc
7141 ctaactgtcc accagtgtct ctgcctgtct ccctgtgtct gcttctcggg ttctctgtgc
7201 catggtggct ctggctacct gtccatcagt gtctccattt ctgttcctcc ccctcagggt
7261 gactctgggg gacccctggt gtgcagagga cagctccagg gcctcgtgtc ttggggaatg
7321 gagcgctgcg ccctgcctgg ctaccccggt gtctacacca acctgtgcaa gtacagaagc
7381 tggattgagg aaacgatgcg ggacaaatga tggtcttcac ggtgggatgg acctcgtcag
7441 ctgcccaggc cctcctctct ctactcagga cccaggagtc caggccccag cccctcctcc
7501 ctcagaccca ggagtccagg cccccagccc ctcctccctc agacccggga gtccaggccc
7561 ccagcccctc ctccctcaga cccaggagtc caggccccag cccctcctcc ctcagacccg
7621 ggagtccagg cccccagccc ctcctccctc agacccagga gtccaggccc cagtccctcc
7681 tccctcagac ccaggagtcc aggcccccag cccctcctcc ctcagaccca ggaatccagg
7741 cccagccccc cctccctcag acccaggagc cccagtcccc cagcccctcc tccttgagac
7801 ccaggagtcc aggcccagcc cctcctccct cagacccagg agccccagtc cccagcatcc
7861 tgatctttac tccggctctg atctctcctt tcccagagca gttgcttcag gcgttttctc
7921 cccaccaagc ccccaccctt gctgtgtcac catcactact caagaccgga ggcacagagg
7981 gcaggagcac agacccctta aaccggcatt gtattccaaa gacgacaatt tttaacacgc
8041 ttagtgtctc taaaaaccga ataaataatg acaataaaaa tggaatcatc ctaaattgta
8101 ttcattcatc catgtgttta ctttttattt tttgagacaa ggtcttgctc agtctcctgg
8161 tgaaatgctg taacgcaatc atagctcact gcaaccgtga cctcctgggc tccagtgatc
8221 ctcttacctc agcctcccga gtagctggga ccacaggtgc ccgtcaccat gccccgctac
//
```

FIGURE 20

The New Human Kallikrein Gene Locus (19q13.3-q13.4) – 300kb